The Journal of Chemical Physics

DFTB+, a software package for efficient approximate density functional theory based atomistic simulations

Cite as: J. Chem. Phys. **152**, 124101 (2020); https://doi.org/10.1063/1.5143190
Submitted: 20 December 2019 . Accepted: 27 February 2020 . Published Online: 23 March 2020

B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshaye, T. Dumitrică, A. Dominguez, S. Ehlert, M. Elstner, T. van der Heide, J. Hermann, S. Irle, J. J. Kranz, C. Köhler, T. Kowalczyk, T. Kubař, I. S. Lee, V. Lutsker, R. J. Maurer, S. K. Min, I. Mitchell, C. Negre, T. A. Niehaus, A. M. N. Niklasson, A. J. Page, A. Pecchia, G. Penazzi, M. P. Persson, J. Řezáč, C. G. Sánchez, M. Sternberg, M. Stöhr, F. Stuckenberg, A. Tkatchenko, V. W.-z. Yu, and T. Frauenheim

COLLECTIONS

Note: This article is part of the JCP Special Topic on Electronic Structure Software.

F This paper was selected as Featured

View Online Export Citation CrossMark

Lock-in Amplifiers
Find out more today

Zurich Instruments

© 2020 Author(s).

The Journal of Chemical Physics | ARTICLE | scitation.org/journal/jcp

DFTB+, a software package for efficient approximate density functional theory based atomistic simulations

Cite as: J. Chem. Phys. 152, 124101 (2020); doi: 10.1063/1.5143190
Submitted: 20 December 2019 • Accepted: 27 February 2020 •
Published Online: 23 March 2020

View Online Export Citation CrossMark

B. Hourahine,[1] B. Aradi,[2,a)] V. Blum,[3] F. Bonafé,[4] A. Buccheri,[5] C. Camacho,[6] C. Cevallos,[6] M. Y. Deshaye,[7] T. Dumitrică,[8] A. Dominguez,[2,9] S. Ehlert,[10] M. Elstner,[11] T. van der Heide,[2] J. Hermann,[12] S. Irle,[13] J. J. Kranz,[11] C. Köhler,[2] T. Kowalczyk,[7] T. Kubař,[11] I. S. Lee,[14] V. Lutsker,[15] R. J. Maurer,[16] S. K. Min,[14] I. Mitchell,[17] C. Negre,[18] T. A. Niehaus,[19] A. M. N. Niklasson,[18] A. J. Page,[20] A. Pecchia,[21] G. Penazzi,[2] M. P. Persson,[22] J. Řezáč,[23] C. G. Sánchez,[24] M. Sternberg,[25] M. Stöhr,[26] F. Stuckenberg,[2] A. Tkatchenko,[26] V. W.-z. Yu,[3] and T. Frauenheim[2,9]

AFFILIATIONS

[1] SUPA, Department of Physics, The University of Strathclyde, Glasgow G4 0NG, United Kingdom
[2] Bremen Center for Computational Materials Science, University of Bremen, Bremen, Germany
[3] Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708, USA
[4] Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany
[5] School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, United Kingdom
[6] School of Chemistry, University of Costa Rica, San José 11501-2060, Costa Rica
[7] Department of Chemistry and Advanced Materials Science and Engineering Center, Western Washington University, Bellingham, Washington 98225, USA
[8] Department of Mechanical Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA
[9] Computational Science Research Center (CSRC) Beijing and Computational Science Applied Research (CSAR) Institute Shenzhen, Shenzhen, China
[10] University of Bonn, Bonn, Germany
[11] Institute of Physical Chemistry, Karlsruhe Institute of Technology, Karlsruhe, Germany
[12] Freie Universität Berlin, Berlin, Germany
[13] Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA
[14] Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan, South Korea
[15] Institut I – Theoretische Physik, University of Regensburg, Regensburg, Germany
[16] Department of Chemistry, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, United Kingdom
[17] Center for Multidimensional Carbon Materials, Institute for Basic Science (IBS), Ulsan 44919, South Korea
[18] Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
[19] Université de Lyon, Université Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, F-69622 Villeurbanne, France
[20] School of Environmental and Life Sciences, University of Newcastle, Callaghan, Australia
[21] CNR-ISMN, Via Salaria km 29.300, 00015 Monterotondo Stazione, Rome, Italy
[22] Dassault Systemes, Cambridge, United Kingdom
[23] Institute of Organic Chemistry and Biochemistry AS CR, Prague, Czech Republic
[24] Instituto Interdisciplinario de Ciencias Básicas, Universidad Nacional de Cuyo, CONICET, Facultad de Ciencias Exactas y Naturales, Mendoza, Argentina

© Author(s) 2020

[25]Argonne National Laboratory, Lemont, Illinois 60439, USA
[26]Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg City, Luxembourg

Note: This article is part of the JCP Special Topic on Electronic Structure Software.
[a)]**Author to whom correspondence should be addressed:** aradi@uni-bremen.de

ABSTRACT

DFTB+ is a versatile community developed open source software package offering fast and efficient methods for carrying out atomistic quantum mechanical simulations. By implementing various methods approximating density functional theory (DFT), such as the density functional based tight binding (DFTB) and the extended tight binding method, it enables simulations of large systems and long timescales with reasonable accuracy while being considerably faster for typical simulations than the respective *ab initio* methods. Based on the DFTB framework, it additionally offers approximated versions of various DFT extensions including hybrid functionals, time dependent formalism for treating excited systems, electron transport using non-equilibrium Green's functions, and many more. DFTB+ can be used as a user-friendly standalone application in addition to being embedded into other software packages as a library or acting as a calculation-server accessed by socket communication. We give an overview of the recently developed capabilities of the DFTB+ code, demonstrating with a few use case examples, discuss the strengths and weaknesses of the various features, and also discuss on-going developments and possible future perspectives.

© 2020 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/1.5143190

I. INTRODUCTION

Density Functional Theory (DFT)[1,2] dominates the landscape of electronic structure methods, being the usual go-to technique to model large, chemically complex systems at good accuracy. For larger systems and time scales, force-field models instead dominate materials and chemical modeling. Between these is the domain of semi-empirical methods, derived from approximations to Hartree–Fock or DFT based methods. Within this space, density functional based tight binding (DFTB)[3–5] effectively offers a reduced complexity DFT method, being derived from a simplification of Kohn–Sham DFT to a tight binding form.[6]

This paper describes the DFTB+ code,[7] an open source implementation, which aims at collecting the developments of this family of methods and making them generally available to the chemical, materials, and condensed matter communities. This article describes extensions to this code since its original release in 2007,[8] there being a lack of a more recent overview of its features and underlying theory.

II. DFTB+ FEATURES

A. The core DFTB-model

The basic DFTB-equations are presented below. They can be easily generalized for periodic cases (k-points) as well as for other boundary conditions, as implemented in DFTB+. All equations throughout this paper are given in atomic units with Hartree as the energy unit.

1. Expansion of the total energy

The DFTB models are derived from Kohn–Sham (KS) DFT[2] by expansion of the total energy functional. Starting from a properly chosen reference density ρ_0 (e.g., superposition of neutral atomic densities), the ground state density is then represented by this reference, as perturbed by density fluctuations: $\rho(\mathbf{r}) = \rho_0(\mathbf{r}) + \delta\rho(\mathbf{r})$. The total energy expression then expands the energy functional in a Taylor series up to third order,

$$E^{\mathrm{DFTB3}}[\rho_0 + \delta\rho] = E^0[\rho_0] + E^1[\rho_0, \delta\rho] + E^2[\rho_0, (\delta\rho)^2] + E^3[\rho_0, (\delta\rho)^3] \tag{1}$$

with

$$\begin{aligned} E^0[\rho_0] &= \frac{1}{2}\sum_{AB}\frac{Z_A Z_B}{R_{AB}} - \frac{1}{2}\iint\frac{\rho_0(\mathbf{r})\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \\ &\quad - \int V^{\mathrm{xc}}[\rho_0]\rho_0(\mathbf{r})\mathrm{d}\mathbf{r} + E^{\mathrm{xc}}[\rho_0], \\ E^1[\rho_0, \delta\rho] &= \sum_i n_i\langle\psi_i|\hat{H}[\rho_0]|\psi_i\rangle, \\ E^2[\rho_0, (\delta\rho)^2] &= \frac{1}{2}\iint\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta^2 E^{\mathrm{xc}}[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\bigg|_{\rho_0}\right)\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}', \\ E^3[\rho_0, (\delta\rho)^3] &= \frac{1}{6}\iiint\frac{\delta^3 E^{\mathrm{xc}}[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')\delta\rho(\mathbf{r}'')}\bigg|_{\rho_0} \\ &\quad \times\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')\delta\rho(\mathbf{r}'')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\mathrm{d}\mathbf{r}'', \end{aligned} \tag{2}$$

with XC being the exchange correlation energy and potential. Several DFTB models have been implemented, starting from the first order non-self-consistent DFTB1[3,4] [originally called DFTB or non-SCC DFTB], the second order DFTB2 (originally called SCC-DFTB),[5] and the more recent extension to third order, DFTB3.[9–12]

2. DFTB1

The first order DFTB1 method is based on three major approximations: (i) it takes only $E^0[\rho_0]$ and $E^1[\rho_0, \delta\rho]$ from Eq. (2) into account, (ii) it is based on a valence-only minimal basis set (ϕ_μ) within a linear combination of atomic orbitals (LCAO) *ansatz*,

$$\psi_i = \sum_\mu c_{\mu i}\phi_\mu, \tag{3}$$

© Author(s) 2020

for the orbitals ψ_i, and (iii) it applies a two-center approximation to the hamiltonian operator $\hat{H}[\rho_0]$.

a. Minimal atomic basis set. The atomic orbital basis set ϕ_μ is explicitly computed from DFT by solving the atomic Kohn–Sham equations with an additional (usually harmonic) confining potential,

$$\left[-\frac{1}{2}\nabla^2 + V^{\text{eff}}[\rho^{\text{atom}}] + \left(\frac{r}{r_0}\right)^n\right]\phi_\mu = \epsilon_\mu\phi_\mu. \tag{4}$$

This leads to slightly compressed atomic-like orbitals for describing the density in bonding situations. The actual values for r_0 are usually given in the publications describing the specific parameterization. The operator $\hat{H}[\rho^0]$ also depends on the superposition of atomic densities, ρ_A (or potentials, V_A^{eff}) of neutral atoms, $\{A\}$, in the geometry being modeled. This density is usually determined from the same atomic KS equations using a slightly different confinement radius, r_0^{d}.

b. DFTB matrix elements. The hamiltonian can be represented in an LCAO basis as

$$H^0_{\mu\nu} = \langle\phi_\mu|\hat{H}[\rho_0]|\phi_\nu\rangle \approx \langle\phi_\mu|-\frac{1}{2}\nabla^2 + V[\rho_A + \rho_B]|\phi_\nu\rangle, \quad \mu \in A, \nu \in B, \tag{5}$$

where the neglect of the three center terms and pseudo-potential contributions[6] lead to a representation, which can be easily computed by evaluating the Kohn–Sham equations for dimers. These matrix elements are computed once as a function of inter-atomic distance for all element pairs. The Slater–Koster[13] combination rules are applied for the actual orientation of these "dimers" within a molecule or solid.

c. Total energy. $E^0[\rho_0]$ depends only on the reference density, so is universal in the sense that it does not specifically depend on the chemical environment (which would determine any charge transfer (CT), $\delta\rho$, occurring). It can, therefore, be determined for a "reference system" and then applied to other environments. This is the key to transferability of the parameters. In DFTB, $E^0[\rho_0]$ is approximated as a sum of pair potentials called repulsive energy terms,

$$E^0[\rho_0] \approx E_{\text{rep}} = \frac{1}{2}\sum_{AB} V^{\text{rep}}_{AB} \tag{6}$$

(see Ref. 14), which are either determined by comparison with DFT calculations[4] or fitted to empirical data.[15] Forces are calculated with the Hellmann–Feynman theorem and derivatives of the repulsive energy.

3. DFTB2 and DFTB3

To approximate the E^2 and E^3 terms in Eq. (2), the density fluctuations are written as a superposition of atomic contributions, taken to be exponentially decaying spherically symmetric charge densities

$$\delta\rho(\mathbf{r}) = \sum_A \delta\rho_A(\mathbf{r} - \mathbf{R}_A) \approx \frac{1}{\sqrt{4\pi}}\sum_A\left(\frac{\tau_A^3}{8\pi}e^{-\tau_A|\mathbf{r}-\mathbf{R}_A|}\right)\Delta q_A. \tag{7}$$

By neglecting the XC-contributions for the moment, the second order integral E^2 leads to an analytical function, γ_{AB}, with energy,[5]

$$E^2(\tau_A, \tau_B, R_{AB}) = \frac{1}{2}\sum_{AB(\neq A)} \gamma_{AB}(\tau_A, \tau_B, R_{AB})\Delta q_A \Delta q_B. \tag{8}$$

The energy depends on the Mulliken charges $\{q_A\}$ (where the atomic charge fluctuation, $\Delta q_A = q_A - Z_A$, is with respect to the neutral atom), which are, in turn, dependent on the molecular orbital coefficients, $c_{\mu i}$. Thus, the resulting equations have to be solved self-consistently. At large distances, γ_{AB} approaches $1/R_{AB}$, while at short distances, it represents electron–electron interactions within one atom. For the limit $R_{AB} \to 0$, one finds $\tau_A = \frac{16}{5}U_A$, i.e., the so-called Hubbard parameter U_A (twice the chemical hardness) is inversely proportional to the width of the atomic charge density τ_A. This relation is intuitive in that more diffuse atoms (or anions) have a smaller chemical hardness. For DFTB, the chemical hardness is computed from DFT, not fitted.

The third order terms describe the change of the chemical hardness of an atom and are also computed from DFT. A function Γ_{AB} results as the derivative of the γ-function with respect to charge, and the DFTB3 total energy is then given by

$$E^{\text{DFTB3}} = \sum_i\sum_{AB}\sum_{\mu\in A}\sum_{\nu\in B} n_i c_{\mu i} c_{\nu i} H^0_{\mu\nu} + \frac{1}{2}\sum_{AB}\Delta q_A \Delta q_B \gamma^h_{AB} + \frac{1}{3}\sum_{AB}\Delta q_A^2 \Delta q_B \Gamma_{AB} + \frac{1}{2}\sum_{AB} V^{\text{rep}}_{AB}. \tag{9}$$

The third order terms become important when local densities deviate significantly from the reference, i.e., Δq_A is large. Apart from including the third order terms, DFTB3 also modifies γ_{AB} for the interactions between hydrogen and first row elements,[9] where the deviation from the relation between the charge width and the chemical hardness, as formulated above, is most pronounced.

The resulting DFTB3 hamiltonian takes the form

$$H_{\mu\nu} = H^0_{\mu\nu} + H^2_{\mu\nu}[\gamma^h, \Delta q] + H^3_{\mu\nu}[\Gamma, \Delta q], \qquad \mu \in A, \nu \in B, \tag{10}$$

$$H^2_{\mu\nu} = \frac{S_{\mu\nu}}{2}\sum_C\left(\gamma^h_{BC} + \gamma^h_{AC}\right)\Delta q_C, \tag{11}$$

$$H^3_{\mu\nu} = S_{\mu\nu}\sum_C\left(\frac{\Delta q_A \Gamma_{AC}}{3} + \frac{\Delta q_B \Gamma_{BC}}{3} + (\Gamma_{AC} + \Gamma_{BC})\frac{\Delta q_c}{6}\right)\Delta q_C, \tag{12}$$

where $S_{\mu\nu}$ is the overlap matrix between orbitals ϕ_μ and ϕ_ν, and γ^h is the modified DFTB2 interaction.

4. Spin

Analogous to DFTB2, expanding the energy with respect to spin fluctuations[16–18] leads to the spin-polarized expressions for DFTB. By introducing the magnetization density $m(\mathbf{r}) = \rho^\uparrow(\mathbf{r}) - \rho^\downarrow(\mathbf{r})$ as difference of the densities of spin-up and spin-down electrons and its corresponding fluctuations $[\delta m(\mathbf{r})]$ around the spin-unpolarized reference state $[|m(\mathbf{r})| = 0]$, a spin dependent term is added to the spin-independent E^2 of Eq. (2),

$$E^2[\rho_0, (\delta\rho)^2, (\delta m)^2] = E^2[\rho_0, (\delta\rho)^2] + \frac{1}{2}\int \left.\frac{\delta^2 E^{\text{xc}}[\rho, m]}{\delta m(\mathbf{r})^2}\right|_{\rho_0, m=0} \times \delta m(\mathbf{r})^2 d\mathbf{r}, \tag{13}$$

© Author(s) 2020

where a local or semi-local E^{xc} has been assumed.

Identifying the spin density fluctuations with up- and down-spin Mulliken charge differences, Δp_{Al}, for angular momentum shell l at atom A, and approximating the second derivative of $E^{xc}[\rho, m]$ as an atomic constant $W_{All'}$ (similar to the Hubbard U_A) lead to an on-site energy contribution

$$E^2_{spin} = \frac{1}{2}\sum_A \sum_{l\in A} \sum_{l'\in A} W_{All'} \Delta p_{Al} \Delta p_{Al'}. \tag{14}$$

This term in Eq. (14) is to be added to Eq. (8). It captures the spin-polarization contribution to the total energy and couples different atomic angular momentum shells via magnetic interaction. The $W_{All'}$ are usually an order of magnitude less than the U_A and are multiplied with a (typically) small Δp_{Al}; hence, inclusion of spin-polarization via Eq. (14) gives only a small energy contribution. If there is a net imbalance of up- and down-spin electrons in the system, the occupation of electronic states alone carries most of the effect of the unpaired electron(s) without including Eq. (14). The use of Mulliken charges leads to an additional hamiltonian contribution[17] to the (now) shell resolved form of Eq. (10),

$$H^{spin\pm}_{\mu\nu} = \pm \frac{S_{\mu\nu}}{2}\left(\sum_{l''\in A} W_{All''}\Delta p_{Al''} + \sum_{l''\in B} W_{Bl'l''}\Delta p_{Bl''}\right), \quad \mu \in l \in A, \nu \in l' \in B, \tag{15}$$

where the spin up (down) hamiltonian has this term added (subtracted).

Expanding further to local (not global) up and down spin populations via Pauli spinors gives the non-collinear spin model.[19] Equation (14) becomes

$$E^2_{spin} = \frac{1}{2}\sum_A \sum_{l\in A} \sum_{l'\in A} W_{All'} \Delta\vec{p}_{Al} \cdot \Delta\vec{p}_{Al'}, \tag{16}$$

and the wave-function generalizes to two component spinors. The hamiltonian contributions take the form

$$\left(H^0_{\mu\nu} + H^2_{\mu\nu} + H^3_{\mu\nu}\right) \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sum_{i=1}^{3} H^{\sigma_i}_{\mu\nu} \otimes \sigma_i, \tag{17}$$

where σ_i is the Pauli matrix for spin component $i(=x, y, z)$ and H^{σ_i} is constructed from the ith spin component of $\Delta\vec{p}$. This spin-block two component hamiltonian then also enables spin–orbit coupling[19,20] to be included in DFTB+. The spin-block hamiltonian addition is

$$H^{L\cdot S}_{\mu\nu} = \frac{S_{\mu\nu}}{2} \otimes \left(\xi_{Al}\begin{pmatrix} L_z & L^- \\ L^+ & -L_z \end{pmatrix}_l + \xi_{Bl'}\begin{pmatrix} L_z & L^- \\ L^+ & -L_z \end{pmatrix}_{l'}\right), \quad \mu \in l \in A, \nu \in l' \in B, \tag{18}$$

where ξ_{Al} is the spin orbit coupling constant for shell l of atom A with $L^\pm$ and L_z being the angular momentum operators for atomic shells.

5. Limitations of the core DFTB-model

DFTB is an approximate method, and as such shows limitations, which can be traced back to the different approximations applied. However, the fitting of Eq. (6) can compensate for some of the inaccuracies. Since until now, only bonding contributions are addressed by the two-center nature of the repulsive potentials, bond-lengths, bond-stretch frequencies, and bond-energies can be targeted (properties such as bond angles or dihedral angles cannot be influenced by repulsive pair parameterization). This is the reason why DFTB performs better than a fixed minimal basis DFT method, which would be only of limited use in most of the applications. In some cases, DFTB can even perform better than double-zeta (DZ) DFT using generalized gradient approximation (GGA) functionals, as shown, e.g., in Ref. 12. This accuracy definitely can be traced back to the parameterization.

a. Integral approximations. There are some approximations in DFTB that cannot be compensated by parameterization, effecting, e.g., bond angles and dihedrals, which on average show an accuracy slightly less than DFT/DZ. Furthermore, the integral approximation leads to an imbalanced description of bonds with different bond order. For example, C–O single, double, and triple bonds have to be covered by a single repulsive potential, which shows only a limited transferability over the three bonding situations. This is the reason why both good atomization energies and vibrational frequencies cannot be covered with a single fit.[12] Hence, in that work, two parameterizations were proposed, one for obtaining accurate energies and one for the vibrational frequencies. Similarly, description of different crystal phases with the same chemical composition but with very different coordination numbers can be challenging. Recent examples show,[21,22] however, that it is possible to reach a reasonable accuracy if special care is taken during the parameterization process.

b. Minimal basis set. The minimal basis set used has several clear limitations, which show up in the overall DFTB performance: First, for a good description of hydrogen in different bonding situations, relatively diffuse wave functions have to be chosen. For this atomic wave-function, however, the H_2 atomization energy is in error, which is dealt with by an *ad hoc* solution, again providing a special repulsive parameter set.[12] Furthermore, nitrogen hybridization and proton affinities require at least the inclusion of d-orbitals in the basis set: this again can be compensated by a special parameter set, which has to be applied under certain conditions.[12] A similar problem occurs for highly coordinated phosphorus containing species.[23] The minimal basis can also become problematic when describing the high lying (conduction band) states in solids. For example, silicon needs d-orbitals in order to describe the conduction band minimum properly. The valence band, on the other hand, can be reasonably described with an sp-only basis.

c. Basis set confinement. As a result of the orbital confinement, Pauli repulsion forces are underestimated, which leads to DFTB non-bonding interactions being on average too short by 10%–15%. This has been investigated in detail for liquid water, where a different repulsive potential has been suggested.[24] A related problem concerns molecular polarizabilities, which are underestimated using a minimal basis set. Approaches to correct for this shortcoming have been summarized recently in Ref. 25. The too-confined range of basis functions also impairs the calculation of electron-transfer couplings. Here, unconfined basis sets have to be used.[26] Similarly, it can be challenging to find a good compromise for the basis confinement when describing 2D-layered materials. As the interlayer distances are significantly longer than the intra-layer ones, the binding between the layers often becomes weaker compared to DFT.

© Author(s) 2020

d. DFT inherited weaknesses. DFTB is derived from DFT and uses standard DFT functionals, which also come with some well-known limitations. There, several strategies applied within DFT are also viable for DFTB, as discussed below in more detail.

B. Density matrix functionals

The typical behavior of the SCC-DFTB ground state resembles local-density approximation (LDA) or GGA,[27] i.e., a mean-field (MF) electronic structure method with associated self-interaction errors and, for some systems, qualitatively incorrect ground states. This is in contrast to non-SCC DFTB, which gives the correct linearity of total energy and step-wise chemical potentials[28] for fractionally charged systems. However, non-SCC can also produce MF-DFT limits, such as in the case of dimer dissociation[29,30] due to self-interaction errors in the underlying atomic DFT potentials.

DFTB+ now also supports long-range corrected hybrid functionals for exchange and correlation. With respect to conventional local/semi-local functionals, these are known to provide a better description of wave function localization and significantly reduce self-interaction.[31] In the longer term, DFTB+ will continue to develop post-DFT based methods with the aim of making large ($\gtrsim$1000 atom) correlated systems tractable via methods with correlated self-energies or wave-functions.

1. Onsite corrections

DFTB2 neglects on-site hamiltonian integrals of the type $(\mu\nu|\mu\nu)$, where ϕ_μ and ϕ_ν are two *different* atomic orbitals of the same atom [both Eq. (5) and the use of Mulliken charges give on-site elements only for $\delta_{\mu\nu} = 1$]. A generalized *dual* population[32] can be introduced as

$$Q^{A,l}_{\mu\nu} = \frac{1}{2}\sum_{\kappa}(\rho_{\mu\kappa}S_{\kappa\nu} + S_{\mu\kappa}\rho_{\kappa\nu}), \qquad l \in A;\ \mu,\nu \in l, \tag{19}$$

where $Q^{A,l}_{\mu\nu}$ is a population matrix for shell l of atom A and the diagonal of each block represents the conventional Mulliken charges for orbitals in the lth shell. Based on this population, all fluctuations of the atomic parts of the density matrix from the reference can be included, not only the diagonal (charge) elements. These must then be treated self-consistently during the calculation. This generalization leads, for example, to an improved description of hydrogen bonds in neutral, protonated, and hydroxide water clusters as well as other water-containing complexes.[33]

The onsite-corrected DFTB requires additional atomic parameters; these are not tunable but computed numerically using DFT (see Ref. 34 for details of their evaluation). The onsite parameter for some chemical elements can be found in the DFTB+ manual. The calculation requires convergence in the dual density populations. This is a somewhat heavier convergence criterion than just charge convergence, and thus, the computational time is moderately affected.

2. DFTB+U and mean-field correlation corrections

For correlated materials such as NiO, a popular correction choice in DFT is the LDA+U family of methods,[35] which add a contribution to the energy of the specified local orbitals obtained from the Hubbard model. The rotationally invariant[36] form of LDA+U can be written in terms of several choices of local projections of the density matrix.[32] Likewise, the double-counting between the Hubbard-model and the density functional mean-field functional take several limiting cases.[37] In DFTB+, the fully localized limit of this functional was implemented early in the code's history[27] using the populations of Eq. (19). Originally applied for rare-earth systems,[38] DFTB+U gives excellent agreement with GGA+U.[39] A closely related correction, pseudo-SIC,[40] where the local part of the self-interaction is removed, modifying only the occupied orbitals, is also available. These approximations lower the energies of occupied atomic orbitals within the specified atomic shells with the aim of removing self-interaction or more accurately representing self-energy. However, as with its use in DFT, this approximation suffers from three main drawbacks. First, the form of the correction depends on the choice of double counting removal.[41] The correlation is also mean-field in nature; hence, all equally filled orbitals within a shell receive the same correction, and therefore, cases not well described by a single determinant are not systematically improved. Finally, the choice of the U (and J) values is not necessarily obvious, with a number of different empirical, linear response, and self-consistent choices possible. Specific to DFTB,[42] the U values may also require co-optimization with the repulsive parameters, in particular, for systems where the electronic structure is geometrically sensitive.

3. Long-range corrected hybrid functionals

a. Single determinant formulation. To correct longer range errors, the electron–electron interactions can be split into short and long range components based on a single parameter ω,

$$\frac{1}{r} = \frac{\exp(-\omega r)}{r} + \frac{(1-\exp(-\omega r))}{r}. \tag{20}$$

The short range contribution is treated in a local or semi-local density functional approximation, while the long range term gives rise to a Hartree–Fock-like exchange term in the hamiltonian.[31] The necessary adaptions for the DFTB method (termed LC-DFTB) were introduced in Refs. 43 and 44. Note that quite generally for DFTB+, the exchange-correlation functional is effectively chosen by loading the appropriate Slater–Koster files created for the desired level of theory. This also holds for LC-DFTB, where different values for the range-separation parameter, ω, lead to different Slater–Koster files. The database at `www.dftb.org` currently hosts the ob2 set[45] for the elements O, N, C, and H with $\omega = 0.3\ a_0^{-1}$.

LC-DFTB calculations can also be performed for spin-polarized systems, enabling evaluation of triplet excited states and their corresponding relaxed geometries. It also paves the way for a rational determination and tuning[31] of the range-separation parameter ω, which amounts to total energy evaluations for neutral and singly ionized species. Note that the required atomic spin constants are functional specific. The spin parameters for the ob2 Slater–Koster set are available in the manual.

b. Spin restricted ensemble references. Instead of single determinants, the spin-restricted ensemble-referenced Kohn–Sham (REKS) method and its state-interaction state-averaged variant (SI-SA-REKS or SSR)[46–51] based on ensemble density functional theory are now available in DFTB+. SSR can describe electronic states with multi-reference character and can accurately calculate excitation

© Author(s) 2020

energies between them (see Sec. II C 2). The SSR method is formulated in the context of the LC-DFTB method (LC-DFTB/SSR)[52] since a long-range corrected functional is crucial to correctly describe the electronic structure particularly for the excited states (see Ref. 52 for details of the formalism). Spin-polarization parameters are also required to describe open-shell microstates. It was observed that LC-DFTB/SSR sometimes gives different stability of the open-shell singlet microstates from the conventional SSR results, depending on excitation characters. In such a case, a simple scaling of atomic spin constants is helpful to account for correct excitation energies (see Ref. 52 for the required scaling of spin constants). The LC-DFTB/SSR method can be extended in the future by using larger active spaces or with additional corrections such as the onsite or DFTB3 terms.

4. Non-covalent interactions

In large systems, non-covalent interactions (van der Waals/vdW forces) between molecules and between individual parts of structures become of key importance. The computational performance of DFTB makes these systems accessible, but large errors are observed for these weaker interactions. Being derived from (semi-)local density-functional theory, DFTB naturally shares the shortcomings of these approximations. This includes the lack of long-range electron correlation that translates to underestimated or missing London dispersion. An accurate account of vdW forces is essential in order to reliably describe a wide range of systems in biology, chemistry, and materials science. DFTB has already been successfully combined with a range of different correction schemes[53–58] to account for these weaker interactions, but here we outline some newer methods available in DFTB+.

a. H5 correction for hydrogen bonds. The H5 correction[59] addresses the issue of hydrogen bonding at the level of the electronic structure. For DFTB2 and DFTB3, interaction energies of H-bonds are severely underestimated for two main reasons: most importantly, the monopole approximation does not allow on-atom polarization; even if this limitation is lifted, the use of minimal basis does not allow polarization of hydrogen. In the H5 correction, the gamma function (Sec. II A 3) is multiplied by an empirical term enhancing the interactions at hydrogen bonding distances between hydrogen atoms and electronegative elements (N, O, and S). The H5 correction is applied within the SCC cycle, thus including many-body effects (the source of the important cooperativity in H-bond networks). The H5 correction was developed for DFTB3 with the 3OB parameters and a specific version of the DFT-D3[60,61] dispersion correction. Note that this D3 correction also includes an additional term augmenting hydrogen–hydrogen repulsion at short range (necessary for an accurate description of aliphatic hydrocarbons[62,63]).

b. DFT-D4 dispersion correction. The D4 model[64,65] is now available in DFTB+ as a dispersion correction. Like D3, pairwise C_6^{AB} dispersion coefficients are obtained from a Casimir–Polder integration of effective atomic polarizabilities $\alpha_{A/B}^{\mathrm{eff}}(iu)$,

$$C_6^{AB} = \frac{3}{\pi}\int_0^\infty \alpha_A^{\mathrm{eff}}(iu)\alpha_B^{\mathrm{eff}}(iu)\mathrm{d}u. \tag{21}$$

The influence of the chemical environment is captured by using a range of reference surroundings, weighted by a coordination number. D4 improves on its predecessor by also including a charge scaling based on atomic partial charges determined as either Mulliken[64] or classical electronegativity equilibration.[65] Especially for metal-containing systems, the introduced charge dependence improves thermochemical properties.[66] Large improvements can also be observed for solid-state polarizabilities of inorganic salts.[67] For a full discussion on the methodology behind D4, we refer the reader to Ref. 65, and the implementation details are presented in Ref. 67. The damping parameters for several Slater–Koster sets are provided in the supplementary material.

To investigate the performance of the DFTB-D4 parameterizations, we evaluate the association energies for the S30L benchmark set.[68,69] DFTB-D4 is compared to DFTB3(3ob)-D3(BJ),[54] GFN1-xTB,[70] and GFN2-xTB;[71] additionally, we include the dispersion corrected SCAN[72] functional in comparison to DFT. The deviation from the reference values is shown in Fig. 1. For the mio parameterization, complexes 4, 15, and 16 were excluded due to missing Slater–Koster parameters. The direct comparison of DFTB3(3ob)-D3(BJ) with a MAD of 7.1 kcal/mol to the respective D4 corrected method with a MAD of 6.5 kcal/mol shows a significant improvement over its predecessor. The DFTB2(mio)-D4 gives an improved description with a MAD of 4.5 kcal/mol, which is better than GFN1-xTB with a MAD of 5.5 kcal/mol. The best performance is reached with GFN2-xTB due to the anisotropic electrostatics

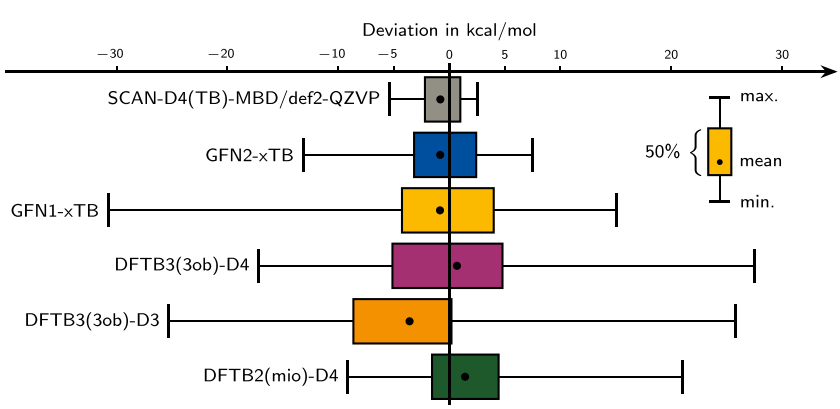

FIG. 1. Performance of different dispersion corrected tight binding methods on the S30L benchmark set, and the values for SCAN-D4 are taken from Ref. 65.

© Author(s) 2020

and the density dependent D4 dispersion, giving a MAD of 3.6 kcal/mol.

c. Tkatchenko–Scheffler (TS) dispersion. The Tkatchenko–Scheffler (TS)[73] correction includes vdW interactions as London-type atom-pairwise C_6/R^6-potentials with damping at short interatomic separations, where the electronic structure method already captures electron correlation. The suggested damping parameters for the mio and 3ob parameter sets are listed in the supplementary material. In the TS approach, all vdW parameters including the static atomic dipole polarizability, α, and C_6-dispersion coefficients depend on the local electronic structure and the chemical environment.[73] High-accuracy *in vacuo* reference values (labeled by vac) are rescaled via

$$x^2 = \left(\frac{\alpha_A^{\mathrm{eff}}}{\alpha_A^{\mathrm{vac}}}\right)^2 = \frac{C_{6,\mathrm{eff}}^{AA}}{C_{6,\mathrm{vac}}^{AA}}. \tag{22}$$

In the case of DFT, x is approximated based on the Hirshfeld atomic volumes.[74] When combined with DFTB, a fast yet accurate alternative has been proposed,[58] which does not require evaluating a real-space representation of the electron density. Instead, the ratio between atom-in-molecule and *in vacuo* net atomic electron populations [i.e., $\mathrm{tr}(\rho)_A/Z_A$] is used to define x.

d. Many-body dispersion (MBD). Going beyond pairwise interactions, many-body dispersion (MBD)[75,76] accounts for many-atom interactions in a dipolar approximation up to infinite order in perturbation theory. This is achieved by describing the system as a set of coupled polarizable dipoles[75] with rescaled *in vacuo* reference polarizabilities [as in Eq. (22)]. At short-ranges, this model switches, via a Fermi-like function with a range of β, to the local atomic environment as accounted for by solving a Dyson-like self-consistent screening equation.[76] β represents a measure for the range of dynamic correlation captured by the underlying electronic structure method, so it depends on the density functional or DFTB parameterization. The recommended β-values for the mio and 3ob parameter sets are listed in the supplementary material.

Figure 2 and Ref. 58 demonstrate that DFTB and MBD represent a promising framework to accurately study long-range correlation forces and emergent behavior at larger length- and timescales. Recently, the DFTB+MBD approach has allowed the study of organic molecular crystals[55] and solvated biomolecules, revealing the complex implications of many-body vdW forces for proteins and their interaction with aqueous environments.[82] Further improvements of TS and MBD, including a better description of charge transfer effects[83] and variational self-consistency,[84] may also be incorporated into DFTB in the future. Both methods are formulated independently of the underlying electronic-structure methods. As a result, DFTB+ outsources the evaluation of the MBD and TS interactions to Libmbd,[85] an external open-source library.

C. Excited states and property calculations

1. Time dependent DFTB with Casida formalism

Electronic excited states are accessible in DFTB+ through time dependent DFTB methods (see Ref. 86 for a review and detailed discussion of this formalism). In a linear response treatment in the frequency domain, excitation energies are obtained by solving an eigenvalue problem known as Casida or RPA (random phase approximation) equations. Compared to first-principles time dependent DFT, the computational scaling can be reduced in DFTB from N^6 to N^3. This is due to the Mulliken approximation for two-electron integrals,[87] which uses transition charges $q_A^{pq\sigma}$,

$$q_A^{pq\sigma} = \frac{1}{4}\sum_{\mu\in A}\sum_{\nu}\left(c_{\mu p}^{\sigma}\tilde{c}_{\nu q}^{\sigma} + c_{\mu q}^{\sigma}\tilde{c}_{\nu p}^{\sigma} + c_{\nu p}^{\sigma}\tilde{c}_{\mu q}^{\sigma} + c_{\nu q}^{\sigma}\tilde{c}_{\mu p}^{\sigma}\right), \quad \tilde{\mathbf{c}}_p = \mathbf{c}_p \cdot \mathbf{S}, \tag{23}$$

for transitions from the Kohn–Sham orbital $p\sigma$ to $q\sigma$.

For fixed geometry, DFTB+ provides a user defined number of low lying excitation energies, oscillator strengths, and orbital participations. In another mode of operation, the code computes excited state charges, eigenvectors of the Casida equation, and energy gradients for a specific state of interest, which can be combined with

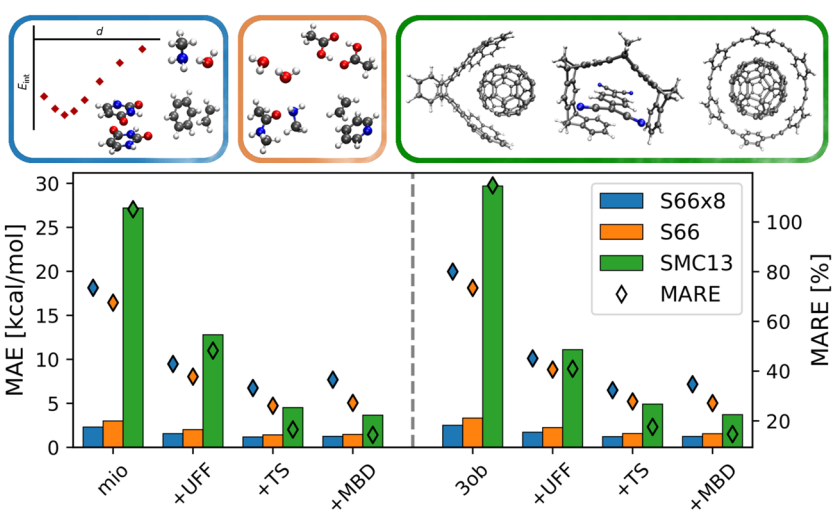

FIG. 2. Mean absolute errors (MAEs) and mean absolute relative errors (MAREs) in inter-molecular interaction energies of bare DFTB and with different van der Waals models in comparison to high-level reference data. S66 and S66x8: small organic dimers and their dissociation curves,[77,78] SMC13: set of 13 supra-molecular complexes.[79–81]

© Author(s) 2020

MD or geometry relaxation. For spin-unpolarized calculations, the response matrix is block diagonal for the singlet and triplet channels to speed up the computation. DFTB+ allows for the computation of the excited state properties of systems with general fractional occupation of the KS orbitals. This is useful, for example, for the simulations of metals and semi-metals at a finite temperature. For a detailed discussion on spin-polarization and fractional occupation within time dependent (TD) DFTB, see Ref. 34. The onsite correction, discussed in Sec. II B 1, is also possible for excited state calculations and was shown to lead to marked improvements.[34]

Due to their improved treatment of charge-transfer transitions, range-separated functionals are also relevant in the context of excited states. DFTB+ implements the time dependent long range corrected (TD-LC) DFTB method, as described in Ref. 88. Compared to the conventional TD-DFTB, the lower symmetry of the response matrix leads to a non-Hermitian eigenvalue problem, which we solve by the algorithm of Stratmann and co-workers.[89] Somewhat surprisingly, it turns out that TD-LC-DFTB calculations are, in practice, not significantly slower than TD-DFTB calculations (see Ref. 88 for a deeper discussion). Gradients can also be calculated with TD-LC-DFTB, making it possible to perform geometry optimizations and MD simulations in singlet excited states.

Note that energetically high lying states and Rydberg excitations are clearly outside of the scope of TD(-LC)-DFTB since their description generally requires very diffuse basis sets. Apart from this class, the photochemically more relevant set of low energy valence excitations are predicted with similar accuracy to first principles TD-DFT, as several benchmarks indicate.[34,90,91] As mentioned above, charge-transfer excitations can now be also treated using TD-LC-DFTB.[88]

2. SSR and excitations

Currently, the SSR method implemented in DFTB+ is formulated for active spaces including two electrons in two fractionally occupied orbitals [i.e., SSR(2,2)], which is suitable for a singlet ground state and the lowest singlet excited state as well as a doubly excited state.[52] In addition, since the SSR method is based on an ensemble representation and includes the electronic correlation, it can give correct state-interactions among nearly degenerate electronic states. Thus, the SSR approach is useful to investigate conical intersections. The LC-DFTB/SSR method with scaled spin constants can accurately describe the ground and excited states including π/π^* or n/π^* character, undergoing bond cleavage/bond formation reactions as well as the conical intersections where the conventional (TD)DFTB fails to obtain the electronic properties. Analytic energy gradients as well as non-adiabatic couplings are also available.

3. Time-independent excited states from ΔDFTB

The linear response approach to excited-state properties in DFTB is efficient and powerful, but there exist circumstances where a more direct route to the excited states is desirable. For example, the excited-state properties obtained from linear response theory require an additional order of derivatives relative to the ground state. As noted in Sec. II C 1, linear-response TD-DFTB (like its parent method TD-DFT)[92] should invoke range-separation to achieve a qualitatively correct picture of charge-transfer excitations and related long-range phenomena.[88]

As an alternative to the time-dependent linear-response approach, it is possible to variationally optimize certain electronically excited states directly. The ΔDFTB method, modeled on the Δ-self-consistent-field (ΔSCF) approach to excited states in DFT,[93,94] involves solving the SCC-DFTB equations subject to an orbital occupation constraint that forces the adoption of a non-aufbau electronic configuration consistent with the target excited state. This method is implemented for the lowest-lying singlet excited state of closed-shell molecules in DFTB+.[95] The converged, non-aufbau SCC-DFTB determinant is a spin-contaminated or "mixed" spin state, but the excitation energy can be approximately spin-purified through the Ziegler sum rule, which extracts the energy of a pure singlet from the energies of the mixed state and the triplet ground state.

A significant advantage of the ΔDFTB approach is that excited-state gradients and hessians are quite straightforward to compute, both mathematically and in terms of computational cost, relative to linear response approaches. Benchmarks of ΔDFTB excited-state geometries and Stokes shifts[95] demonstrate the suitability of the method for simulating excited-state energetics and dynamics of common organic chromophores along the S_1 potential energy surface.

4. Real-time propagation of electrons and Ehrenfest dynamics

It is often desirable to study time dependent properties outside the linear response regime, e.g., under strong external fields. The numerical propagation of the electronic states enables the simulation of such phenomena, and its coupling to the nuclear dynamics in a semi-classical level can be included to the lowest order within the Ehrenfest method. Purely electronic (frozen-nuclei) dynamics as well as Ehrenfest dynamics are included in DFTB+. We solve the equation of motion of the reduced density matrix ρ given by the Liouville-von Neumann equation

$$\dot{\rho} = -\mathrm{i}\left(S^{-1}H[\rho]\rho - \rho H[\rho]S^{-1}\right) - \left(S^{-1}D\rho + \rho D^{\dagger}S^{-1}\right), \tag{24}$$

with D being the non-adiabatic coupling matrix $D_{\mu\nu} = \dot{\mathbf{R}}_B \cdot \nabla_B S_{\mu\nu}$ and $\dot{\mathbf{R}}_B$ being the velocity of atom B. The on-site blocks can be calculated taking the $\mathbf{R}_B \to 0$ limit, although neglecting those does not introduce significant changes to the dynamics.[96]

Unitary evolution of ρ with no change in its eigenvalues would require $D^{\dagger} = -D$, which is normally not the case. Therefore, nuclear dynamics can induce electronic transitions leading to thermalization.[97] Unitary evolution is recovered when all nuclear velocities are equal (frozen-nuclei dynamics) and the second term in Eq. (24) vanishes.

The force in the Ehrenfest-dynamics can be expressed as[96,98]

$$\begin{aligned}\mathbf{F}_A = &-\mathrm{Tr}\left\{\rho\left(\nabla_A H^0 + \nabla_A S\sum_B \gamma_{AB}\Delta q_B + \nabla_A S S^{-1}H + HS^{-1}\nabla_A S\right)\right\}\\ &- \mathrm{i}\,\mathrm{Tr}\left\{\rho\nabla_A S S^{-1}D + \mathrm{h.c.}\right\} + \mathrm{i}\sum_{\mu\nu}\left\{\rho_{\nu\mu}\langle\nabla_A\phi_\mu|\nabla_B\phi_\nu\rangle\cdot\dot{\mathbf{R}}_B + \mathrm{h.c.}\right\}\\ &- \Delta q_A\sum_B \nabla_A\gamma_{AB}\Delta q_B - \nabla_A E_{rep} - \Delta q_A\mathbf{E}(t),\end{aligned} \tag{25}$$

where $\mathbf{E}(t)$ is the external electric field. In the present implementation, the velocity dependent terms have been neglected, and they would vanish for a complete basis[96] and are necessary for momentum, but not for energy conservation.[98] When the system is driven

© Author(s) 2020

externally by an electric field, a dipole coupling term is added in the time-dependent hamiltonian in Eq. (24).

Some applications that have been enabled by the speedup over time-dependent DFT are the simulation of the plasmon-driven breathing-mode excitation in silver nanoparticles of 1–2 nm in diameter[99] and the simulation of transient absorption pump–probe spectra in molecules.[100,101]

Whenever a time propagation approach is used for the calculation of absorption spectra in the linear regime, this method is equivalent to calculations using the Casida formalism and shares its strengths and limitations. Specific pitfalls of the time dependent approach come into play whenever simulating the response to intense external fields. In these cases, the poor description of highly lying excited states due to the use of a minimal basis set would likely be inaccurate if these states are populated during the dynamics.

5. pp-RPA

An approximate particle–particle RPA scheme, the so-called pp-DFTB,[88] is now implemented in DFTB+. Particle–particle RPA, based on the pairing matrix fluctuation formalism, has been shown to be an efficient approach for the accurate description of double and charge-transfer (CT) excitations involving the highest occupied molecular orbital (HOMO) (see Ref. 102 for details). In Ref. 88, we compare against TD-LC-DFTB for CT excitation energies of donor–acceptor complexes. TD-LC-DFTB has the advantage that transitions do not necessarily have to involve the HOMO of the system. Alternatively, pp-DFTB does not require parameter tuning and is efficient for the lowest lying excitations.

Although one of the strengths of the original pp-RPA formulation lies on the accurate description of Rydberg excitations, our approximate formalism based on DFTB fails to describe these kinds of transitions, as explained in Sec. II C 1.

6. Coupled perturbed responses

DFTB+ supports several types of response calculations for second-order derivatives. The general form of the response evaluation is via standard perturbation theory,

$$P_{ij} = \left\langle c_i \middle| H_{ij}^{(1)} - \epsilon_j S_{ij}^{(1)} \middle| c_j \right\rangle, \tag{26}$$

$$\epsilon_i^{(1)} = P_{ij}\delta_{ij}, \tag{27}$$

$$U_{ij} = P_{ij}/(\epsilon_j - \epsilon_i), \tag{28}$$

$$c_i^{(1)} = \sum_j U_{ij} c_j^{(0)}, \tag{29}$$

$$\rho^{(1)} = \sum_i n_i^{(1)} \left|c^{(0)}\right\rangle\left\langle c^{(0)}\right| + \sum_i n_i^{(0)} \left(\left|c^{(1)}\right\rangle\left\langle c^{(0)}\right| + \text{c.c.}\right), \tag{30}$$

where the sums for the states that U mixes together may be over all states or only the virtual space (parallel gauge) depending on application. U is anti-symmetric (anti-Hermitian) or has no symmetry depending on whether the derivative of the overlap matrix is non-zero.

In the case of systems with degenerate levels, a unitary transformation, Z, that diagonalizes the block of P associated with that manifold can be applied to the states; note that this sub-block is always symmetric (Hermitian), leading to orthogonality between states in the perturbation operation,

$$\tilde{P}_{ij} = z_{ik} P_{kl} z_{li}^{\dagger}, \tag{31}$$

$$\tilde{c}_i = c_j z_{ji}. \tag{32}$$

For fractionally occupied levels, the derivatives of the occupations for $\mathbf{q} = 0$ perturbations (where the change in the Fermi energy should be included) are then evaluated.[103]

Time dependent perturbations at an energy of $\hbar\omega$ can be written as

$$U_{ij}^{\pm} = P_{ij}/(\epsilon_j - \epsilon_i \pm \hbar\omega + i\eta), \tag{33}$$

$$c_i^{(1)\pm} = \sum_j U_{ij}^{\pm} c_j^{(0)}, \tag{34}$$

$$\rho^{(1)} = \sum_i n_i^{(1)} \left|c^{(0)}\right\rangle\left\langle c^{(0)}\right| + \sum_i n_i^{(1)} \left|c^{(0)}\right\rangle\left\langle c^{(0)}\right| + \sum_{\pm}\sum_i n_i^{(0)} \left(\left|c^{(1)\pm}\right\rangle\left\langle c^{(0)}\right\rangle + \text{c.c.}\right). \tag{35}$$

Here, the small constant η prevents divergence exactly at excitation poles.

Derivatives with respect to external electric fields and potentials are included (giving polarizabilities and dipole excitation energies), with respect to atom positions (at $\mathbf{q} = 0$, providing Born charges and electronic derivatives for the hessian) and with respect to k in periodic systems (effective masses and also the Berry connection via $\langle u|\partial u/\partial \mathbf{k}\rangle$). In the longer term, perturbation with respect to magnetic fields, boundary conditions (elastic tensors), and alternative approaches (Sternheimer equations for $\mathbf{q} \neq 0$, and also lower computationally scaling density matrix perturbation theory) are planned.

D. Non-equilibrium Green's function based electron transport

Electron transport in the steady-state regime is described in DFTB+ within a non-equilibrium Green's function (NEGF) method,[104,105] as implemented in the code-independent libNEGF[106] library. The density matrix is evaluated in terms of the electron–electron correlation matrix $G^{<}$,[105]

$$\rho = \frac{1}{2\pi \mathrm{i}} \int_{-\infty}^{+\infty} G^{<}(E)\mathrm{d}E. \tag{36}$$

Open boundary conditions are included in terms of electron baths with an arbitrary spectrum and chemical potential, allowing for a seamless description of charge injection from electrodes with an applied bias. The density matrix is then used to evaluate a real-space electron density distribution, which is coupled self-consistently with a Poisson solver. We perform a full band integration of Eq. (36), utilizing a complex contour integral to reduce the number of integration points.[104] This allows for an implicit description of dielectric properties, which is crucial for an accurate modeling of ultra-scaled electron devices.[107,108] After self-consistency is achieved, the total

© Author(s) 2020

current flowing in the system is calculated with the Landauer/Caroli formula for the non-interacting case or with the Meir–Wingreen formula for the interacting case.[105] A detailed description of the numerical algorithms and self-consistent coupling is presented in Ref. 109. Here, we summarize the main features that might differentiate DFTB+ from other nano-device simulation packages: (i) support for $N \geqslant 1$ electrodes (enabling structures from surfaces and semi-infinite wires to multiple terminal geometries), (ii) $O(L)$ memory and time scaling (where L is the system length) via a block-iterative algorithm, (iii) a real space Poisson solver with support for gates and dielectric regions, and (iv) evaluation of local currents. Being a parameterized tight binding method, its usage is bounded by the availability of good parameters for the system under investigation.

Carbon-based materials and molecular junctions have been a typical use-case since the first integration of DFTB and NEGF.[110–112] In Fig. 3, we show a non-SCC calculation example of transmission in linear response for a multi-terminal device. The simulated system is a cross-junction between two (10,10) Carbon nanotubes (CNTs). One CNT is tilted by 60° with respect to the second, and the transmission is calculated by displacing one CNT along the axis of the other. The transmission follows, as expected, a periodic pattern in accordance with the lattice repeat of 0.25 nm along the axis of the CNT.

Currently, we are working on extending transport functionality in DFTB+ with electron–phonon coupling,[113–116] electron–photon coupling, spin polarized transport, and phonon transport.[117–120]

Overall, DFTB-NEGF shares many similarities with DFT based implementations, and it also inherits some shortcomings the less experienced users should be aware of. For example, the open boundary treatment demands that external and non-equilibrium potentials are screened at the boundaries.[105] Therefore, the simulated system should be large enough compared to the screening length. This condition is easily achieved with bulk metallic electrodes, but it can be difficult with low dimensional systems that exhibit poor screening. When this condition is not fulfilled, unphysical discontinuities in the potential may be obtained. In addition, compared to band structure calculations, NEGF tends to converge with more difficulty.[121] Aside from these common challenges, it is important that for DFTB-NEGF calculations, any set of parameters should be evaluated by verifying at the least band structure properties in the energy range of interest. DFTB parameters fitted to reproduce total energies and forces might be excellent in those applications but lack the necessary accuracy in the band structure. Depending on the degree of accuracy required, an *ad hoc* fitting for transport calculations could also be necessary, for example, in the case of silicon.[122]

E. Extended Lagrangian Born–Oppenheimer dynamics

The Extended Lagrangian Born–Oppenheimer molecular dynamics (XLBOMD) framework allows[123,124] molecular dynamics on the Born–Oppenheimer surface with only one hamiltonian diagonalization per time step without the need for self-consistency cycles. The basic idea is based on a backward error analysis, i.e., instead of calculating approximate forces through an expensive non-linear iterative optimization procedure for an underlying exact potential energy surface, XL-BOMD calculates exact forces for an approximate "shadow" potential energy surface, $U(\mathbf{R}, n)$. This is approximated from a constrained minimization of a linearized Kohn–Sham energy functional.[124,125] The functional is linearized around an approximate ground state density, n. This density is included as a dynamical field variable driven by an extended harmonic oscillator centered on an approximate ground state, $q[n]$, which is given by the minimization of the linearized Kohn–Sham functional. The harmonic well is defined in terms of a metric tensor, $T = K^T K$, where the kernel K is assumed to be the inverse Jacobian of the residual function, $q[n] - n$.[124] The equations of motion are given by

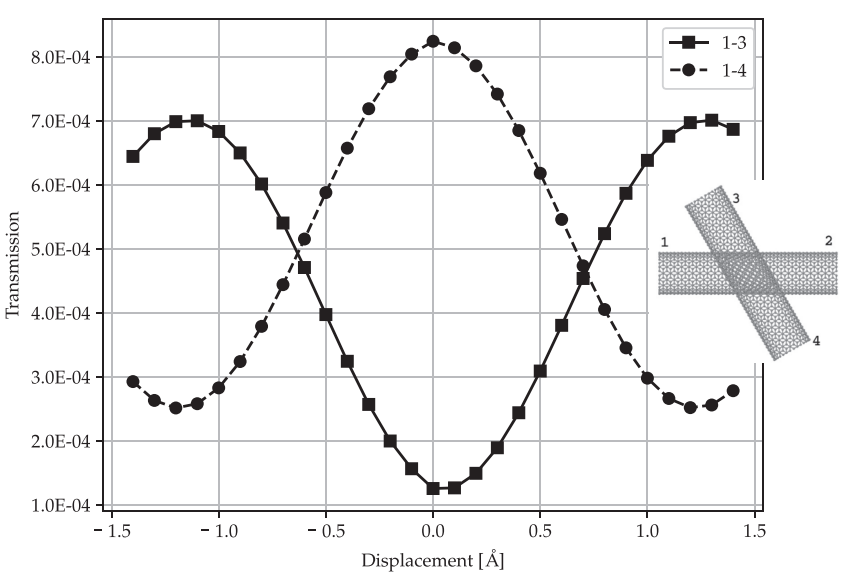

FIG. 3. Transmission across two (10,10) CNTs as a function of the displacement of the top CNT along the axis of the bottom CNT. The two curves represent the transmission resolved between electrode 1 of the bottom CNT, and, respectively, electrodes 2 and 3 of the top CNT (as labeled in the inset).

© Author(s) 2020

$$M_I\ddot{\mathbf{R}}_I = -\left.\frac{\partial U(\mathbf{R},n)}{\partial R_I}\right|_n \quad \text{and} \quad \ddot{n} = -\omega^2 K(q[n]-n). \tag{37}$$

Here, M_I are the atomic masses, $\mathbf{R}_I$ are the nuclear coordinates, ω is the frequency of the harmonic oscillator, $q[n]$ are the net Mulliken charge vectors (from an optimized linearized energy expression), and n is the extended dynamical variable that is set to the optimized ground state net Mulliken charge vector in the initial time step. The details of the DFTB + implementation are given in Ref. 126.

We currently approximate the kernel by a scaled identity matrix,

$$K = -cI, \quad c \in [0,1]. \tag{38}$$

For many problems, this is a sufficiently accurate approximation. However, for the most challenging problems including simulations of reactive chemical systems or metals, the scaled delta function is not a sufficiently stable approximation. Improved approximations have been developed[124] and will be implemented in the DFTB+ program in the near future.

F. Objective geometries

Objective structures[127] (OSs) describe geometries consisting of a set of identical cells, where the corresponding atoms in different cells can be mapped onto each other by orthogonal transformation(s). Both finite and infinite OSs are possible. Currently, we describe structures[127–129] possessing C_n rotational symmetry and a $C_m \otimes T$ screw axis, where $n \in \mathbb{N}^*$ and $m \in \mathbb{R}^+$,

$$\mathbf{X}_{i,\zeta,\xi} = (C_n)^{\xi}(C_m)^{\zeta}\mathbf{X}_i + T^{\zeta}, \quad i \in N, \tag{39}$$

with N atoms in the reference objective cell ($\{\mathbf{X}_i\}$) and $\{\zeta,\xi\} \in \mathbb{N}$, where $-\infty < \zeta < \infty$ and $0 < \xi < n$. Exploiting the objective boundary conditions (OBCs) can introduce substantial computational savings, for example, irrational values of m lead to structures with a small OS cell, but an infinitely long one dimensional periodic boundary condition (PBC), i.e., intractable purely as a T operation. OBCs generalize symmetry-adapted Bloch sums for orbitals. As with molecular and periodic structures,[8] most expressions in DFTB+ can be performed in real space via the boundary-condition agnostic and sparse representation of matrices in real space, the solution of the hamiltonian only requires dense matrices and k-points. For the long-range Coulombic and dispersion interactions in DFTB, we also require lattice sums that are generalized to these boundary conditions.[130]

Further examples can be found in Refs. 131–133, but here we demonstrate the bending of a BN bi-layer. Figure 4 shows a double-walled tubular OS with a curvature of $1/R$ (from the tube radius) that represents the bent bi-layer. Bending along the **a** (**b**) direction of the sheet is an "armchair" ("zig-zag") tube with a C_n proper axis, described as an eight atom objective cell in which we select $\mathbf{T} = \mathbf{a}$ ($\mathbf{T} = \mathbf{b}$), with no tube twist. The bi-layer bends as a plate, with the outer wall stretching and the inner wall compressing along their circumferential directions; its energy change is interpreted as bending strain (E_{bend}). It is important to note that the corresponding curvature is not an imposed constraint, but a result of the calculation: R is the average tube radius. Figure 4(b) demonstrates linearity with bending, and fitting to $E_{\text{bend}} = (1/2)D(|\mathbf{a}||\mathbf{b}|)(1/R)^2$ gives a bi-layer bending constant of $D = 120$ eV.

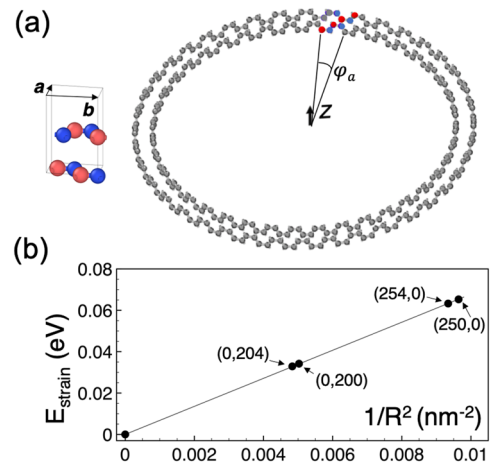

FIG. 4. (a) OS of a BN bi-layer tube with a B_4N_4 unit (red and blue atoms). Angular, but not translational, objective images are shown in gray. (b) Bending energy (circles) vs curvature with a linear fit.

A wider range of OSs will be made available in later DFTB+ releases, along with adapted electrostatic evaluation for these structures.

G. Extended tight binding hamiltonian

The extended tight binding (xTB) methods were primarily designed for the fast calculation of structures and non-covalent interaction energies for finite systems with a few thousand atoms. The main parameterizations, GFNn-xTB, target molecular geometries, frequencies, and non-covalent interactions follow mostly a global and element-specific parameter only strategy. The historically first parameterization, GFN1-xTB, covers all elements up to $Z = 86$ and is now supported in DFTB+. Its successor, GFN2-xTB,[71] will also be made available in the future.

We briefly outline the xTB methods; for a more detailed discussion and comparison to other methods, we refer to Refs. 70 and 71. The xTB core hamiltonian is constructed in a partially polarized STO-nG basis set with diagonal terms made flexible by adding a dependence on the local chemical environment according to a coordination number (CN), similar to that used in DFT-D3,[60]

$$H_{\lambda\lambda} = H^l_A - H^l_{CN_A} CN_A. \tag{40}$$

The off-diagonal terms are approximated as an average of the diagonal terms proportional to the overlap between the corresponding basis functions.

Both GFN1-xTB and GFN2-xTB include density fluctuation up to third order diagonal terms, while the distance dependence of the Coulomb interaction within the isotropic second order term is described by a generalized form of the Mataga–Nishimoto–Ohno–Klopman[134–136] expression. In GFN2-xTB, the expansion of the second order density fluctuations goes beyond the usual isotropic

© Author(s) 2020

energy terms and includes interactions up to R^{-3}, i.e., charge–dipole, dipole–dipole, and charge–quadrupole interactions, which significantly improves the description of inter-molecular interactions, such as halogen bonds and hydrogen bonds, without the need to include force-field-like corrections as in DFTB or GFN1-xTB. It is planned to implement full multipole electrostatics with Ewald summation in DFTB+ to enable GFN2-xTB and other generalized DFTB models.[137]

GFN1-xTB and GFN2-xTB have been extensively tested for their target properties,[71] and further studies regarding structures for lanthanoid complexes[138] and transition metal complexes[66] have shown xTB methods to be robust for all its parameterized elements. Errors in these methods are very systematic, which can be used to devise correction schemes for off-target properties such as reaction enthalpies.[139]

H. DFTB parameterization

1. Parameterization workflow

a. Electronic parameters. The electronic parameterization for DFTB involves two principal steps. First, the compressed atomic densities and the atomic basis functions have to be determined (a one-center problem), followed by the calculation of the hamiltonian and overlap elements at various distances (a two-center problem). The compressed densities and wave-functions come from solving a Kohn–Sham-problem for a single atom with an additional confinement potential (usually a power function), as shown in Eq. (4). One may use different compression radii (and make separate calculations) to obtain the compressed density and the compressed atomic wave-functions for a given atom. The atomic calculations are currently carried out with a code implementing the Hartree–Fock theory based atomic problem[140,141] extended with the possibility of including DFT exchange-correlation potentials via the libxc library[142] and scalar relativistic effects via the zero-order relativistic approximation (ZORA).[143] The resulting densities and atomic wave-functions are stored on a grid. The two-center integration tool reads those grid-based quantities and calculates the hamiltonian and overlap two-center integrals for various distances using the Becke-method.[144]

b. Repulsive parameters. Once the electronic parameters for certain species have been determined, the first three terms of Eq. (9) can be calculated for any systems composed of those species. The missing fourth term, the repulsive energy, is composed of pairwise contributions, V_{AB}^{rep}, between all possible atomic pairs of A and B in the system [see Eq. (6)]. During the parameterization process, one aims to determine repulsive potentials between the atomic species as a function of the distance between the atoms R_{AB} so that $V_{AB}^{\mathrm{rep}} = f_{\mathrm{sp}(A),\mathrm{sp}(B)}(R_{AB})$, where sp($X$) refers to the species of atom X. In contrast to the electronic parameters, which are determined by species-specific parameters only, the repulsive functions must be defined for each combination of species pairs separately. They are usually determined by minimizing the difference between the reference (usually *ab initio*) total energies and the DFTB total energies for a given set of atomic geometries. If one uses only one (or a few simple) reference system(s), the optimal repulsive function can be determined manually, while for more complex scenarios, usually semi-automatic approaches[15,21,145–147] are used.

2. Outlook

In recent years, machine learning has been utilized with DFTB+, usually to enhance the generation and description of the repulsive potentials[148–152] or try to improve on electronic parameters.[151,153] Related Δ-machine learning[154] methodologies based on neural network corrections for DFTB energies and forces have been also reported recently.[155,156] We are currently in the process of developing a new unified machine-learning framework, which for a target system allows optimal adaption of both the electronic and the repulsive contributions. Given the predicted DFTB model, one would still have to solve it in order to obtain the system properties. On the other hand, changing external conditions (temperature, electric field, and applied bias) would not require additional training in this approach, and long range effects (e.g., metallic states) could also be described easily.

III. TECHNICAL ASPECTS OF THE DFTB+ PACKAGE

A. Parallel scaling

In large-scale simulations, the solution of the DFTB hamiltonian to obtain the density matrix eventually becomes prohibitively expensive, scaling cubically with the size of the system being simulated. The diagonalization infrastructure in DFTB+ has undergone a major upgrade, including distributed parallelism and GPU accelerated solutions to address this cost. If instead the density matrix is directly obtained from the hamiltonian, circumventing diagonalization, then linear or quadratic scaling can now be obtained, depending on the chosen method. DFTB+ will continue to benefit from developments in these advanced solvers as we move into the era of exascale computing.

1. The ELSI interface and supported solvers

ELSI[157] features a unified software interface that simplifies the use of various high-performance eigensolvers (ELPA[158] EigenExa,[159] SLEPc,[160] and MAGMA[161]) and density matrix solvers (libOMM,[162] PEXSI,[163] and NTPoly[164]). We convert the sparse DFTB+ H and S structures[8] into either standard 2D block-cyclic distributed dense matrices or sparse 1D block distributed matrices compatible with the ELSI interface. All k-points and spin channels are then solved in parallel.

The ELSI-supported solvers, when applied in appropriate cases, can lead to a substantial speedup over the default distributed parallel diagonalization method in DFTB+, i.e., eigensolvers in the ScaLAPACK library.[165–167] Figure 5 demonstrates two examples: non-self-consistent-charge, spin-non-polarized, Γ-point calculations for a C_{64000} nanotube (CNT) and a Si_{6750} supercell, with 25 600 and 27 000 basis functions, respectively. Figure 5(c) shows the time to build the density matrix for the CNT model with three solvers, the pdsyevr eigensolver in the MKL implementation of ScaLAPACK, the ELPA2 eigensolver, and the PEXSI density matrix solver. Here, both the MKL's version of pdsyevr eigensolver and the ELPA2 eigensolver adopt a two-stage tri-diagonalization algorithm.[158,168,169] In terms of performance, ELPA2 and MKL pdsyevr are similar, while both are outperformed by the PEXSI solver by more than an order of magnitude. The PEXSI[163] method directly constructs the density matrix from the hamiltonian and overlap matrices with a computational

© Author(s) 2020

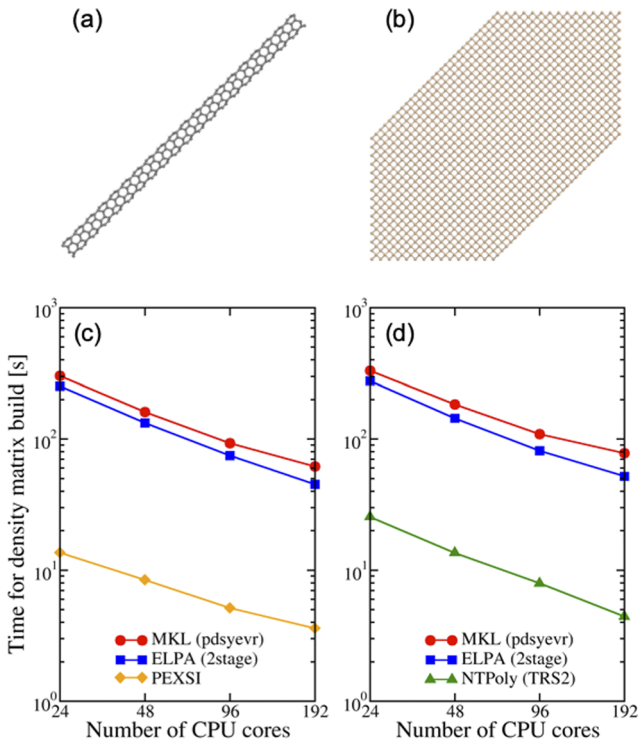

FIG. 5. Atomic structures of (a) the carbon nanotube (CNT) model (6400 atoms) and (b) the silicon supercell model (6750 atoms). The length of the actual CNT model is 16 times that of the structure shown in (a). (c) and (d) show the time to compute the density matrix for models (a) and (b), respectively. Calculations are performed on the NewRiver computer. MKL pdsyevr and ELPA2 first compute all the eigenvalues and eigenvectors of the eigensystem of H and S and then build the density matrix. PEXSI and NTPoly directly construct the density matrix from H and S.

complexity of $O(N^{(d+1)/2})$ for $d = 1 \ldots 3$D systems. This reduced scaling property stems from sparse linear algebra, not the existence of an energy gap. Therefore, for any low-dimensional system, regardless of the electronic structure, PEXSI can be used as a powerful alternative to diagonalization. A similar comparison of solver performance for the silicon supercell model is shown in Fig. 5(d), where the NTPoly density matrix solver shows greater performance than the MKL pdsyevr and ELPA2 eigensolvers. Around its massively parallel sparse matrix multiplication routine, NTPoly implements various linear scaling density matrix purification methods, including the 2nd order trace-resetting purification method (TRS2)[170] used here. While PEXSI is not particularly suited for 3D systems, NTPoly offers an alternative as long as the system has a non-trivial energy gap.

B. Order-N scaling with the SP2 solver

The SP2 (second-order recursive spectral projection expansion),[170] which is valid at zero electronic temperature when $1/k_BT \to \infty$, recursively expands a Heaviside step function to project the (occupied) density matrix,

$$\rho = \lim_{n\to\infty} F_n(F_{n-1}(\ldots F_0(H^{\perp})\ldots)), \tag{41}$$

where $H^{\perp}$ is the hamiltonian transformed into an orthogonalized basis, given by the congruence transformation, $H^{\perp} = Z^T HZ$. Each iteration of the SP2 Fermi-operator expansion consists of a generalized matrix–matrix multiplication that can be performed using thresholded sparse matrix algebra. In this way, the computational complexity in each iteration can be reduced to $O(N)$ for sufficiently large sparse matrices. Note that we cannot expect linear scaling complexity for metals, since the inter-atomic elements of the density matrix decay algebraically instead of exponentially.[171] The spectral projection functions in the SP2 expansion can be chosen to correct $\mathrm{Tr}(\rho)$ such that the step is formed automatically around the chemical potential separating the occupied from the unoccupied states.[170] Obtaining the congruence matrix, Z, introduces a potential $O(N^3)$ bottleneck. To avoid this, the sparsity of S can be exploited and the Z matrix can be obtained recursively with linear scaling complexity applying the "ZSP method" developed in Refs. 172 and 173.

Several versions of the SP2 algorithm can be found in the PROGRESS library,[174] which uses the Basic Matrix Library (BML)[175,176] for the thresholded sparse matrix–matrix operations. The matrix data structure is based on the ELLPACK-R sparse matrix format, which allows efficient shared memory parallelism on a single node.[177] The DFTB+ code was modified to use the LANL PROGRESS library and, in particular, the SP2 and ZSP algorithms. In combination with XL-BOMD, this allows efficient energy-conserving, molecular dynamics simulations, where the computational cost scales only linearly with the system size. Figure 6 shows the performance of the SP2 algorithm compared to regular diagonalization.

C. GPU computing

Graphics processing unit (GPU) acceleration is implemented in DFTB+. Given the nature of the underlying theory, the time-limiting step in routine calculations corresponds to the diagonalization of the hamiltonian matrix, taking in the order of 90%–95% of

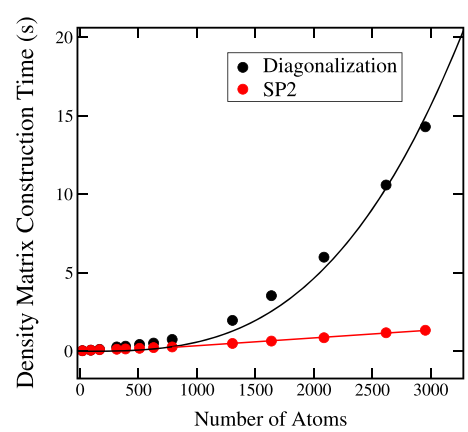

FIG. 6. CPU time for the density matrix construction for different varying sizes of water box systems. Regular diagonalization (black curve) was compared to the SP2 method (red curve). A numerical threshold of 10^{-5} was used in the sparse matrix–matrix multiplications of the SP2 algorithm.

© Author(s) 2020

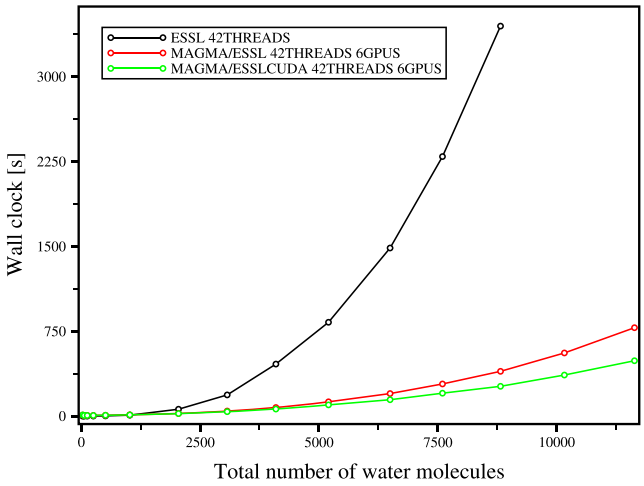

FIG. 7. Wall clock running times for total energy calculations of water clusters (with 6 basis functions/water molecule). The black curve shows timings obtained using the LAPACK compatible ESSL eigensolver on the CPU, and the red/green curves show timings obtained using the MAGMA and the ESSL libraries without/with ESSL-CUDA off-loading. Timings have been recorded on the Summit machine using 42 threads for 42 physical cores.

the total running time. The hybrid CPU–GPU implementation in DFTB+ replaces the LAPACK-based eigensolver with a GPU eigensolver based on the divide-and-conquer algorithm as implemented in MAGMA.[178]

Benchmarking of the code shows that at least 5000 basis functions are necessary to exploit the power of the GPUs and to produce an observable speedup with respect to the CPU-only code. For systems spanning a vector space comprised of 70 000 basis functions, speedups of 17× have been observed in a system with 6 NVIDIA® Tesla® V100 with respect to the multi-threading CPU-only implementation (see Fig. 7).

IV. INTERFACING DFTB+ WITH OTHER SOFTWARE PACKAGES

DFTB+ can be currently interfaced with other software packages using three different ways of communications: file communication, socket based, or direct connection via the DFTB + API as a library. The first one is very easy to implement but comes with an overhead for the file I/O, while the latter two enable a more efficient coupling at the price of somewhat higher complexity in implementation.

A. File based communication

When using file based communication, the external driver creates necessary input files and starts an individual DFTB+ program for each of the inputs. After DFTB+ has finished, the driver analyses the created output files and extracts the necessary information from those. DFTB+ had been interfaced using file based communication to, among others, the phonopy[179] code for finite difference harmonic and anharmonic phonon calculations and the Atomic Simulation Environment (ASE) package[180] (a set of tools and Python modules for setting up, manipulating, running, visualizing, and analyzing atomistic simulations).

B. Socket interface

The i-PI[181] interface for communication with external driving codes is supported by DFTB+. DFTB+ can then be driven directly instead of using file I/O. The initial input to DFTB+ specifies the boundary conditions, type of calculation, and chemical information for atoms, and the code then waits to be externally contacted. This kind of communication with DFTB+ can be used by, among others, the i-PI universal force engine package[181] and ASE.[180]

C. DFTB+ library, QM/MM simulations

1. Gromacs integration

DFTB quantum-chemical models may be utilized as a QM engine in hybrid quantum-mechanical/molecular mechanical (QM/MM) approaches. This allows, for example, efficient simulations of chemical processes taking place in bio-molecular complexes. The DFTB+ library interface has been connected to the Gromacs[182] MM-simulation software package. (The Gromacs part of the integration is contained in a fork of the Gromacs main branch.[183]) At the start of the simulation, the DFTB + input file is read in, and a DFTB calculation environment is created, containing all of the necessary information (parameters), but no atomic coordinates yet. In every step of MD simulation or energy minimization, the calculation of forces starts with a call to the DFTB+ API, passing the coordinates of QM atoms and the values of electrostatic potentials induced by the MM atoms at the positions of the QM atoms. DFTB+ then returns QM forces and QM charges back to Gromacs, where the QM/MM forces are calculated in the QM/MM routines. Gromacs then continues by calculating the MM forces, integration of equations of motion, etc.

Sometimes the electrostatic interactions cannot be represented as an external potential but also depend on the actual values of the QM-charges (i.e., polarizable surroundings). In those cases, a callback function can be passed to DFTB+, which is then invoked at every SCC iteration to update the potential by the driver program whenever the QM charges change. In the DFTB+/Gromacs integration, we use this technique to calculate the QM–QM electrostatic interactions in periodic systems with the highly efficient particle mesh Ewald method[184] implemented in Gromacs.

2. DL_POLY_4 integration with MPI support

DL_POLY_4 is a general-purpose package for classical molecular dynamics (MD) simulations.[185] In conjunction with the recent extension of DFTB+'s API, DL_POLY_4.10 supports the use of DFTB+ for self-consistent force calculations in place of empirical force fields for Born–Oppenheimer molecular dynamics.

The interface fully supports passing MPI communicators between the programs, allowing users to run simulations in parallel, across multiple processes. The MPI parallelization schemes of DL_POLY_4 and DFTB+ differ considerably. DL_POLY_4 utilizes domain decomposition to spatially distribute the atoms that comprise the system across multiple processes, whereas DFTB+ distributes the hamiltonian matrix elements using BLACS

© Author(s) 2020

decomposition. This does not impose any serious restrictions as DL_POLY_4 and DFTB+ run sequentially, with DFTB+ being called once per MD time step.

The DL_POLY_4–DFTB+ interface works by gathering the atoms from each DL_POLY_4 process such that all processes have a complete copy of all the atoms. Coordinates, species types, and the atomic ordering are then passed to DFTB+. The calculated forces are returned to DL_POLY_4, which redistributes them according to its domain decomposition, and the atomic positions are propagated one time step.

Spatial decomposition means that atoms can propagate between processes. Because atoms are gathered sequentially according to their process id (or rank), when atoms propagate between processes, their ordering effectively changes. The DFTB+ API facilitates this and is, therefore, able to support any molecular dynamics software that implements domain decomposition parallelization; however, the total number of atoms (and atom types) must be conserved during the simulation.

D. Meta-dynamics using PLUMED

Molecular dynamics is often plagued by high energy barriers that trap the nuclear ensemble in one or several local minima. This leads to inefficient or inadequate sampling of the ensemble and thus inaccurate predictions of physicochemical properties.[186–188] This "timescale" problem is typical for rare-event systems or those in which ergodicity of a particular state is impeded by the local topology of the potential energy surface. A variety of methods have been conceived to circumvent this, including umbrella sampling[189] and meta-dynamics.[190]

Umbrella sampling and meta-dynamics can now be performed using DFTB+ via its interface to the PLUMED plugin.[191,192] Using PLUMED, MD trajectories generated in DFTB+ can be analyzed, sampled, and biased in a variety of ways along user-defined collective variables (CVs), enabling accelerated MD simulations and determination of the free energy surface. A CV is a subspace of the full potential energy surface that can be arbitrarily defined to sample atomic dynamics along dimensions/pathways of physicochemical interest. PLUMED also includes bias functions such as the upper and lower wall biases, enabling a constraint of MD configurations to specific areas on the potential energy surface. The utility of the DFTB+/PLUMED interface has been demonstrated on several challenging systems, including malonaldehyde intra-molecular proton transfer (Fig. 8), corannulene bowl inversion, and the diffusion of epoxide groups on graphene.[192]

E. DFTB+ in Materials Studio

DFTB+ is included as a module in the commercial modeling and simulation software package, BIOVIA Materials Studio (MS).[193] DFTB+ runs as an in-process energy server, supplying energies, forces, and stresses to drive the MS in-house simulations tools. Supported tasks include energy calculation, geometry optimization, molecular dynamics, electron transport calculation, mechanical properties, and parameterization. The module also supports calculation and visualization of standard electronic properties, such as band structure, density of states, orbitals, and so on. The DFTB+ module integrates closely with the data model and the Materials Studio Visualizer, allowing the user to construct structures and start calculations quickly, with fully automated creation of the DFTB+ input file. The DFTB+ module is also supported in the MS MaterialsScript interface and the Materials Studio Collection for Pipeline Pilot,[194] allowing creation of more complicated workflows.[195,196]

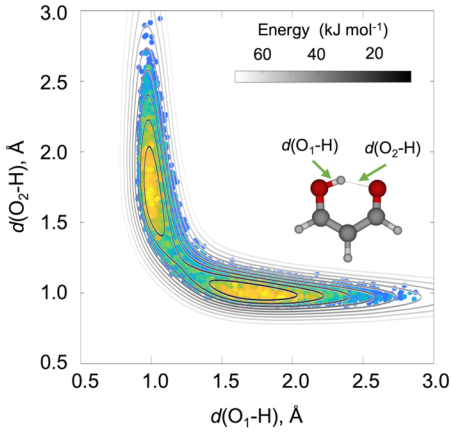

FIG. 8. Intra-molecular proton transfer in malonaldehyde at 298 K. Contours show the DFTB3-D3/3ob free energy surface of malonaldehyde obtained using well-tempered meta-dynamics, with collective variables $d(O_1–H)$ and $d(O_2–H)$. Each point is colored according to its sampling frequency during the meta-dynamics simulation, yellow (blue) indicating high (low) sampling frequency. The DFTB3-D/3ob free energy surface yields a proton transfer barrier of 13.1 ± 0.4 kJ mol^{-1}.

The DFTB+ parameterization workflow in MS supports fitting of both electronic parameters and repulsive pair potentials using DFT calculations with the DMol3 module[197,198] as a reference. The DFTB+ module includes scripts for validation of parameters in terms of band structure, bond length, bond angles, and so on, as well as visualization for the hamiltonian, overlap matrix elements, and the repulsive pair potentials. The parameterization tools allow extension of existing parameters or incremental development of a parameter set. Parameters developed using the DFTB+ module can, after conversion, be used outside MS. Several default DFTB+ parameter sets, generated using these parameterization tools, are also included. In 2019, MS introduced a new parameter set that includes the Li, C, H, N, O, P, and F elements and is aimed toward Li-ion battery modeling.

F. Outlook

In order to enable flexible general communication with various types of external components (external drivers, QM/MM, and machine learning models), we are in the process of developing a communication library,[199] which allows for data exchange between mixed language (e.g., Fortran and C) components via API-bindings as well as between different processes via socket communications. After engagement with other stakeholders, this will be released as a set of BSD-licensed tools and a library.

V. SOFTWARE ENGINEERING IN DFTB+

This section presents a few aspects of our software development, which may have some interest beyond the DFTB+ software package.

© Author(s) 2020

A. Modern Fortran wrappers for MPI and ScaLAPACK functions

Modern scientific modeling packages must be able to run on massive parallel architectures to utilize high performance computing, often using the Message Passing Interface (MPI) framework. While the MPI offers a versatile parallelization framework, its application interface was designed to support C and Fortran 77-like interfaces. This requires the programmer to explicitly pass arguments to the MPI-routines, which *should* be automatically deduced by the compiler for languages with higher abstraction levels (C++ or Fortran 95 and newer versions). In order to eliminate developer need to pass redundant information (and to reduce associated programming bugs), we have developed modern Fortran wrappers around the MPI-routines. These have been collected in the MPIFX-library,[200] which is an independent software project outside of the DFTB+ software suite, being licensed under the more permissive BSD-license. It enables shorter MPI-calls by automatically deducing data types and data sizes from the call signature. Additionally, several MPI parameters have been made optional using their most commonly used value as a default value. For example, in order to broadcast a real array from the master process to all other process, one would have to make the following MPI-call:

```
call mpi_bcast(array, size(array), MPI_FLOAT, 0, comm, err)
```

while MPIFX-wrappers reduce it to a much shorter and less error-prone line:

```
call mpifx_bcast(comm, array)
```

where `comm` is an MPIFX derived type containing the MPI-communicator. The type (`MPI_FLOAT`) and number of broadcasted items [`size(array)`] are automatically deduced. The process initiating the broadcasting has been assumed to be process 0 (master process), as this is probably the most common use case but can be customized when needed with an optional parameter. The error argument is optional as well, if it is not passed (as in the example above), the routine would stop the code in the case of any errors.

Likewise, the commonly used parallel linear algebra library ScaLAPACK uses Fortran 77-type interfaces. The open source SCALAPACKFX library[201] offers higher level modern Fortran wrappers around routines used by DFTB+.

B. Fortran meta-programming using Fypp

Although the latest Fortran standard (Fortran 2018) offers many constructs to support modern programming paradigms, it does not allow for generic template based programming. This would avoid substantial code duplication and offer useful meta-programming capabilities for Fortran programmers. We have developed the Python based pre-processor, Fypp,[202] which offers a workaround for the missing features. Fypp is used during the build process to turn the meta-programming constructs into the standard Fortran code. The Fypp project is independent of the DFTB+ software package and is licensed under the BSD-license, being also used by other scientific software packages, for example, by the CP2K code[203] and both the MPIFX and the SCALAPACKFX libraries.

VI. SUMMARY

DFTB+ is an atomistic quantum mechanical simulation software package allowing fast and efficient simulations of large systems for long timescales. It implements the DFTB- and the xTB-methods and various extensions of those, such as range-separated functionals, multiple methods of excited state calculations, and electron transport simulations. It can be used either as a standalone application or as a library and has been already interfaced to several other simulation packages. DFTB+ is a community developed open source project under the GNU Lesser General Public License, which can be freely used, modified, and extended by everybody.

SUPPLEMENTARY MATERIAL

See the supplementary material for the damping parameters for the D4, the TS, and the MBD dispersion models.

ACKNOWLEDGMENTS

The authors, especially B. Hourahine and B. Aradi, are thankful to Gotthard Seifert for his suggestions and insights into density functional tight binding throughout the development of the DFTB+ code. B. Hourahine acknowledges the EPSRC (Grant No. EP/P015719/1) for financial support. B. Aradi and T. Frauenheim acknowledge the research training group DFG-RTG 2247 (QM3). A. Buccheri acknowledges the EPSRC (Grant No. EP/P022308/1). C. Camacho acknowledges financial support from the Vice-Rectory for research of the University of Costa Rica (Grant No. 115-B9-461) and the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory (ORNL), which is managed by UT-Battelle, LLC, for DOE under Contract No. DE-AC05-00OR22725. S. Irle acknowledges support from the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division, Geoscience Program. M. Y. Deshaye and T. Kowalczyk acknowledge support from a National Science Foundation RUI Award (No. CHE-1664674) and a CAREER Award (No. DMR-1848067). T. Kowalczyk is a Cottrell Scholar of the Research Corporation for Science Advancement. T. Dumitrică was supported by the National Science Foundation (Grant No. CMMI-1332228). R. J. Maurer acknowledges support via a UKRI Future Leaders Fellowship (Grant No. MR/S016023/1). A. M. N. Niklasson and C. Negre acknowledge support from the U.S. Department of Energy Office of Basic Energy Sciences (Grant No. FWP LANLE8AN); the U.S. Department of Energy through the Los Alamos National Laboratory; and the Exascale Computing Project (No. 17-SC-20-SC), a collaborative effort of the U.S. Department of Energy, Office of Science and the National Nuclear Security Administration. T. A. Niehaus would like to thank the Laboratoire d'Excellence iMUST for financial support. M. Stöhr acknowledges financial support from the Fonds National de la Recherche, Luxembourg (AFR Ph.D. Grant No. CNDTEC). A. Tkatchenko was supported by the European Research Council (ERC-CoG BeStMo). V. W.-z. Yu was supported by the National Science Foundation (NSF) under Grant No. 1450280 and a fellowship from the Molecular Sciences Software Institute under NSF Grant No. 1547580.

© Author(s) 2020

REFERENCES

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] G. Seifert, D. Porezag, and T. Frauenheim, Int. J. Quantum Chem. **58**, 185 (1996).
[4] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, and R. Kaschner, Phys. Rev. B **51**, 12947 (1995).
[5] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, Phys. Rev. B **58**, 7260 (1998).
[6] M. Elstner and G. Seifert, Philos. Trans. R. Soc. A **372**, 20120483 (2014).
[7] See https://github.com/dftbplus/dftbplus for DFTB+ software package; accessed 15 December 2019.
[8] B. Aradi, B. Hourahine, and T. Frauenheim, J. Phys. Chem. A **111**, 5678 (2007).
[9] M. Elstner, J. Phys. Chem. A **111**, 5614 (2007).
[10] Y. Yang, H. Yu, D. York, Q. Cui, and M. Elstner, J. Phys. Chem. A **111**, 10861 (2007).
[11] M. Gaus, Q. Cui, and M. Elstner, J. Chem. Theory Comput. **7**, 931 (2011).
[12] M. Gaus, A. Goez, and M. Elstner, J. Chem. Theory Comput. **9**, 338 (2012).
[13] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).
[14] G. Seifert and J.-O. Joswig, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 456 (2012).
[15] M. Gaus, C.-P. Chou, H. Witek, and M. Elstner, J. Phys. Chem. A **113**, 11866 (2009).
[16] T. Frauenheim, G. Seifert, M. Elstner, Z. Hajnal, G. Jungnickel, D. Porezag, S. Suhai, and R. Scholz, Phys. Status Solidi B **217**, 41 (2000).
[17] C. Köhler, G. Seifert, and T. Frauenheim, Chem. Phys. **309**, 23 (2005).
[18] C. Köhler, G. Seifert, U. Gerstmann, M. Elstner, H. Overhof, and T. Frauenheim, Phys. Chem. Chem. Phys. **3**, 5109 (2001).
[19] C. Köhler, T. Frauenheim, B. Hourahine, G. Seifert, and M. Sternberg, J. Phys. Chem. A **111**, 5622 (2007).
[20] B. Hourahine, MRS Proc. **1290**, 46 (2011).
[21] M. Hellström, K. Jorner, M. Bryngelsson, S. E. Huber, J. Kullgren, T. Frauenheim, and P. Broqvist, J. Phys. Chem. C **117**, 17004 (2013).
[22] A. Fihey, C. Hettich, J. Touzeau, F. Maurel, A. Perrier, C. Köhler, B. Aradi, and T. Frauenheim, J. Comput. Chem. **36**, 2075 (2015).
[23] M. Gaus, X. Lu, M. Elstner, and Q. Cui, J. Chem. Theory Comput. **10**, 1518 (2014).
[24] P. Goyal, H.-J. Qian, S. Irle, X. Lu, D. Roston, T. Mori, M. Elstner, and Q. Cui, J. Phys. Chem. B **118**, 11007 (2014).
[25] A. S. Christensen, T. Kubař, Q. Cui, and M. Elstner, Chem. Rev. **116**, 5301 (2016).
[26] A. Kubas, F. Hoffmann, A. Heck, H. Oberhofer, M. Elstner, and J. Blumberger, J. Chem. Phys. **140**, 104105 (2014).
[27] B. Hourahine, S. Sanna, B. Aradi, C. Köhler, T. Niehaus, and T. Frauenheim, J. Phys. Chem. A **111**, 5671 (2007).
[28] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Phys. Rev. Lett. **49**, 1691 (1982).
[29] M. Rapacioli, F. Spiegelman, A. Scemama, and A. Mirtschink, J. Chem. Theory Comput. **7**, 44 (2011).
[30] M. Lundberg, Y. Nishimoto, and S. Irle, Int. J. Quant. Chem. **112**, 1701 (2012).
[31] R. Baer, E. Livshits, and U. Salzner, Annu. Rev. Phys. Chem. **61**, 85 (2010).
[32] M. J. Han, T. Ozaki, and J. Yu, Phys. Rev. B **73**, 045110 (2006).
[33] A. Dominguez, T. Frauenheim, and T. A. Niehaus, J. Phys. Chem. A **119**, 3535 (2015).
[34] A. Dominguez, B. Aradi, T. Frauenheim, V. Lutsker, and T. A. Niehaus, J. Chem. Theory Comput. **9**, 4901 (2013).
[35] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, J. Phys.: Condens. Matter **9**, 767 (1997).
[36] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton, Phys. Rev. B **57**, 1505 (1998).
[37] A. G. Petukhov, I. I. Mazin, L. Chioncel, and A. I. Lichtenstein, Phys. Rev. B **67**, 153106 (2003).
[38] S. Sanna, B. Hourahine, T. Frauenheim, and U. Gerstmann, Phys. Status Solidi C **5**, 2358 (2008).
[39] H. Liu, G. Seifert, and C. Di Valentin, J. Chem. Phys. **150**, 094703 (2019).
[40] A. Filippetti and N. A. Spaldin, Phys. Rev. B **67**, 125109 (2003).
[41] E. R. Ylvisaker, W. E. Pickett, and K. Koepernik, Phys. Rev. B **79**, 035103 (2009).
[42] J. Kullgren, M. J. Wolf, K. Hermansson, C. Köhler, B. Aradi, T. Frauenheim, and P. Broqvist, J. Phys. Chem. C **121**, 4593 (2017).
[43] T. Niehaus and F. Della Sala, Phys. Status Solidi B **249**, 237 (2012).
[44] V. Lutsker, B. Aradi, and T. A. Niehaus, J. Chem. Phys. **143**, 184107 (2015).
[45] V. Q. Vuong, Y. Nishimoto, D. G. Fedorov, B. G. Sumpter, T. A. Niehaus, and S. Irle, J. Chem. Theory Comput. **15**, 3008 (2019).
[46] A. Kazaryan, J. Heuver, and M. Filatov, J. Phys. Chem. A **112**, 12980 (2008).
[47] M. Filatov, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **5**, 146 (2015).
[48] M. Filatov, *Density-functional Methods for Excited States* (Springer, Heidelberg, 2016), pp. 97–124.
[49] M. Filatov and S. Shaik, Chem. Phys. Lett. **304**, 429 (1999).
[50] I. de P. R. Moreira, R. Costa, M. Filatov, and F. Illas, J. Chem. Theory Comput. **3**, 764 (2007).
[51] M. Filatov, F. Liu, and T. J. Martinez, J. Chem. Phys. **147**, 034113 (2017).
[52] I. S. Lee, M. Filatov, and S. K. Min, J. Chem. Theory Comput. **15**, 3021 (2019).
[53] M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, and E. Kaxiras, J. Chem. Phys. **114**, 5149 (2001).
[54] J. G. Brandenburg and S. Grimme, J. Phys. Chem. Lett. **5**, 1785 (2014).
[55] M. Mortazavi, J. G. Brandenburg, R. J. Maurer, and A. Tkatchenko, J. Phys. Chem. Lett. **9**, 399 (2018).
[56] M. Rapacioli, F. Spiegelman, D. Talbi, T. Mineva, A. Goursot, T. Heine, and G. Seifert, J. Chem. Phys. **130**, 244304 (2009).
[57] R. Petraglia, S. N. Steinmann, and C. Corminboeuf, Int. J. Quantum Chem. **115**, 1265 (2015).
[58] M. Stöhr, G. S. Michelitsch, J. C. Tully, K. Reuter, and R. J. Maurer, J. Chem. Phys. **144**, 151101 (2016).
[59] J. Řezáč, J. Chem. Theory Comput. **13**, 4804 (2017).
[60] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).
[61] S. Grimme, S. Ehrlich, and L. Goerigk, J. Comput. Chem. **32**, 1456 (2011).
[62] J. Řezáč and P. Hobza, J. Chem. Theory Comput. **8**, 141 (2012).
[63] B. Vorlová, D. Nachtigallová, J. Jirásková-Vaníčková, H. Ajani, P. Jansa, J. Řezáč, J. Fanfrlík, M. Otyepka, P. Hobza, J. Konvalinka, and M. Lepšík, Eur. J. Med. Chem. **89**, 189 (2015).
[64] E. Caldeweyher, C. Bannwarth, and S. Grimme, J. Chem. Phys. **147**, 034112 (2017).
[65] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme, J. Chem. Phys. **150**, 154122 (2019).
[66] M. Bursch, E. Caldeweyher, A. Hansen, H. Neugebauer, S. Ehlert, and S. Grimme, Acc. Chem. Res. **52**, 258 (2019).
[67] E. Caldeweyher, J.-M. Mewes, S. Ehlert, and S. Grimme, Phys. Chem. Chem. Phys. (in press); chemrxiv:10.26434/chemrxiv.10299428.
[68] R. Sure and S. Grimme, J. Chem. Theory Comput. **11**, 3785–3801 (2015).
[69] J. G. Brandenburg, C. Bannwarth, A. Hansen, and S. Grimme, J. Chem. Phys. **148**, 064104 (2018).
[70] S. Grimme, C. Bannwarth, and P. Shushkov, J. Chem. Theory Comput. **13**, 1989 (2017).
[71] C. Bannwarth, S. Ehlert, and S. Grimme, J. Chem. Theory Comput. **15**, 1652 (2019).
[72] J. Sun, A. Ruzsinszky, and J. P. Perdew, Phys. Rev. Lett. **115**, 036402 (2015).
[73] A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009).
[74] F. L. Hirshfeld, Theor. Chim. Acta **44**, 129 (1977).
[75] A. Tkatchenko, R. A. DiStasio, Jr., R. Car, and M. Scheffler, Phys. Rev. Lett. **108**, 236402 (2012).
[76] A. Ambrosetti, A. M. Reilly, R. A. DiStasio, Jr., and A. Tkatchenko, J. Chem. Phys. **140**, 18A508 (2014).
[77] J. Řezáč, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 3466 (2011).

© Author(s) 2020

[78] J. Řezáč, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 2427 (2011).
[79] A. Ambrosetti, D. Alfè, R. A. DiStasio, Jr., and A. Tkatchenko, J. Phys. Chem. Lett. **5**, 849 (2014).
[80] J. Hermann, D. Alfè, and A. Tkatchenko, Nat. Commun. **8**, 14052 (2017).
[81] M. Stöhr, T. Van Voorhis, and A. Tkatchenko, Chem. Soc. Rev. **48**, 4118 (2019).
[82] M. Stöhr and A. Tkatchenko, Sci. Adv. **5**, eaax0024 (2019).
[83] T. Gould, S. Lebègue, J. G. Ángyán, and T. Bučko, J. Chem. Theory Comput. **12**, 5920 (2016).
[84] N. Ferri, R. A. DiStasio, Jr., A. Ambrosetti, R. Car, and A. Tkatchenko, Phys. Rev. Lett. **114**, 176802 (2015).
[85] J. Hermann, Libmbd software library, https://github.com/jhrmnn/libmbd/; accessed 15 December 2019.
[86] T. A. Niehaus, J. Mol. Struct.: THEOCHEM **914**, 38 (2009).
[87] T. A. Niehaus, S. Suhai, F. Della Sala, P. Lugli, M. Elstner, G. Seifert, and T. Frauenheim, Phys. Rev. B **63**, 085108 (2001).
[88] J. J. Kranz, M. Elstner, B. Aradi, T. Frauenheim, V. Lutsker, A. Dominguez Garcia, and T. A. Niehaus, J. Chem. Theory Comput. **13**, 1737 (2017).
[89] R. E. Stratmann, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **109**, 8218 (1998).
[90] F. Trani, G. Scalmani, G. Zheng, I. Carnimeo, M. J. Frisch, and V. Barone, J. Chem. Theory Comput. **7**, 3304 (2011).
[91] A. Fihey and D. Jacquemin, J. Chem. Theory Comput. **15**, 6267 (2019).
[92] A. Dreuw and M. Head-Gordon, J. Am. Chem. Soc. **126**, 4007 (2004).
[93] T. Ziegler, A. Rauk, and E. J. Baerends, Theor. Chim. Acta **43**, 261 (1977).
[94] T. Kowalczyk, S. R. Yost, and T. Van Voorhis, J. Chem. Phys. **134**, 054128 (2011).
[95] T. Kowalczyk, K. Le, and S. Irle, J. Chem. Theory Comput. **12**, 313 (2016).
[96] T. A. Niehaus, D. Heringer, B. Torralva, and T. Frauenheim, Eur. Phys. J. D **35**, 467 (2005).
[97] Z. Lin and R. E. Allen, J. Phys.: Condens. Matter **21**, 485503 (2009).
[98] T. N. Todorov, J. Phys.: Condens. Matter **13**, 10125 (2001).
[99] F. P. Bonafé, B. Aradi, M. Guan, O. A. Douglas-Gallardo, C. Lian, S. Meng, T. Frauenheim, and C. G. Sánchez, Nanoscale **9**, 12391 (2017).
[100] F. P. Bonafé, F. J. Hernández, B. Aradi, T. Frauenheim, and C. G. Sánchez, J. Phys. Chem. Lett. **9**, 4355 (2018).
[101] F. J. Hernández, F. P. Bonafé, B. Aradi, T. Frauenheim, and C. G. Sánchez, J. Phys. Chem. A **123**, 2065 (2019).
[102] Y. Yang, A. Dominguez, D. Zhang, V. Lutsker, T. A. Niehaus, T. Frauenheim, and W. Yang, J. Chem. Phys. **146**, 124104 (2017).
[103] Y. Nishimoto and S. Irle, Chem. Phys. Lett. **667**, 317 (2017).
[104] A. Pecchia and A. Di Carlo, Rep. Prog. Phys. **67**, 1497 (2004).
[105] H. Haug and A.-P. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductors* (Springer, Berlin; New York, 2008).
[106] See https://github.com/libnegf/libnegf for libNEGF library; accessed 15 December 2019.
[107] S. Markov, G. Penazzi, Y. Kwok, A. Pecchia, B. Aradi, T. Frauenheim, and G. Chen, IEEE Trans. Electron Devices Lett. **36**, 1076 (2015).
[108] Y. Chu, P. Sarangapani, J. Charles, G. Klimeck, and T. Kubis, J. Appl. Phys. **123**, 244501 (2018).
[109] A. Pecchia, G. Penazzi, L. Salvucci, and A. Di Carlo, New J. Phys. **10**, 065022 (2008).
[110] J. R. Reimers, G. C. Solomon, A. Gagliardi, A. Bilic, N. S. Hush, T. Frauenheim, A. Di Carlo, and A. Pecchia, J. Phys. Chem. A **111**, 5692 (2007) [232nd National Meeting of the American-Chemical-Society, San Francisco, CA, 10–14 September 2006].
[111] L. Latessa, A. Pecchia, A. Di Carlo, and P. Lugli, Phys. Rev. B **72**, 035455 (2005).
[112] G. Penazzi, J. M. Carlsson, C. Diedrich, G. Olf, A. Pecchia, and T. Frauenheim, J. Phys. Chem. C **117**, 8020 (2013).
[113] A. Pecchia, A. Di Carlo, A. Gagliardi, S. Sanna, T. Frauenheim, and R. Gutierrez, Nano Lett. **4**, 2109 (2004).
[114] A. Pecchia, G. Romano, A. Gagliardi, T. Frauenheim, and A. Di Carlo, J. Comput. Electron. **6**, 335 (2007).
[115] G. Penazzi, A. Pecchia, V. Gupta, and T. Frauenheim, J. Phys. Chem. C **120**, 16383 (2016).
[116] A. Gagliardi, G. Romano, A. Pecchia, A. Di Carlo, T. Frauenheim, and T. A. Niehaus, New J. Phys. **10** (2008).
[117] L. Medrano Sandonas, R. Gutierrez, A. Pecchia, A. Dianat, and G. Cuniberti, J. Self-Assem. Mol. Electron. **3** (2015).
[118] L. Medrano Sandonas, D. Teich, R. Gutierrez, T. Lorenz, A. Pecchia, G. Seifert, and G. Cuniberti, J. Phys. Chem. C **120**, 18841 (2016).
[119] D. Martinez Gutierrez, A. Di Pierro, A. Pecchia, L. Medrano Sandonas, R. Gutierrez, M. Bernal, B. Mortazavi, G. Cuniberti, G. Saracco, and A. Fina, Nano Res. **12**, 791–799 (2019).
[120] L. Medrano Sandonas, R. Gutierrez, A. Pecchia, A. Croy, and G. Cuniberti, Entropy **21**, 735 (2019).
[121] T. Ozaki, K. Nishio, and H. Kino, Phys. Rev. B **81**, 035116 (2010).
[122] S. Markov, B. Aradi, C. Yam, H. Xie, T. Frauenheim, and G. Chen, IEEE Trans. Comput. **62**, 696 (2015).
[123] A. M. N. Niklasson, Phys. Rev. Lett. **100**, 123004 (2008).
[124] A. M. N. Niklasson, J. Chem. Phys. **147**, 054103 (2017).
[125] A. M. N. Niklasson and M. Cawkwell, J. Chem. Phys. **141**, 164123 (2014).
[126] B. Aradi, A. M. N. Niklasson, and T. Frauenheim, J. Chem. Theory Comput. **11**, 3357 (2015).
[127] R. James, J. Mech. Phys. Solids **54**, 2354 (2006).
[128] T. Dumitrică and R. D. James, J. Mech. Phys. Solids **55**, 2206 (2007).
[129] D.-B. Zhang, M. Hua, and T. Dumitrică, J. Chem. Phys. **128**, 084104 (2008).
[130] I. Nikiforov, B. Hourahine, B. Aradi, T. Frauenheim, and T. Dumitrică, J. Chem. Phys. **139**, 094110 (2013).
[131] I. Nikiforov, B. Hourahine, T. Frauenheim, and T. Dumitrică, J. Phys. Chem. Lett. **5**, 4083 (2014).
[132] H. Xu, G. Drozdov, B. Hourahine, J. G. Park, R. Sweat, T. Frauenheim, and T. Dumitrică, Carbon **143**, 786 (2019).
[133] T. Dumitrică, Carbohydr. Polym. **230**, 115624 (2020).
[134] G. Klopman, J. Am. Chem. Soc. **86**, 4550 (1964).
[135] K. Ohno, Theor. Chim. Acta **2**, 219 (1964).
[136] K. Nishimoto and N. Mataga, Z. Phys. Chem. **12**, 335 (1957).
[137] Z. Bodrog and B. Aradi, Phys. Status Solidi B **249**, 259 (2012).
[138] M. Bursch, A. Hansen, and S. Grimme, Inorg. Chem. **56**, 12485 (2017).
[139] J. C. Kromann, A. Welford, A. S. Christensen, and J. H. Jensen, ACS Omega **3**, 4372 (2018).
[140] C. C. J. Roothan, Rev. Mod. Phys. **32**, 179 (1960).
[141] C. C. J. Roothan, L. M. Sachs, and A. W. Weiss, Rev. Mod. Phys. **32**, 186 (1960).
[142] S. Lehtola, C. Steigemann, M. J. Oliveira, and M. A. Marques, SoftwareX **7**, 1 (2018).
[143] E. van Lenthe, E. J. Baerends, and J. G. Snijders, J. Chem. Phys. **99**, 4597 (1993).
[144] A. D. Becke, J. Chem. Phys. **88**, 2547 (1988).
[145] V. Q. Vuong, J. A. Kuriappan, M. Kubillus, J. Kranz, T. Mast, T. A. Niehaus, S. Irle, and M. Elstner, J. Chem. Theory Comput. **14**, 115 (2018).
[146] See https://github.com/smarkov/skpar for SKPAR software; accessed 15 December 2019.
[147] See https://bitbucket.org/solccp/adpt_core for Automatized DFTB Parameter Toolkit; accessed 15 December 2019.
[148] J. M. Knaup, B. Hourahine, and T. Frauenheim, J. Phys. Chem. A **111**, 5637 (2007).
[149] J. J. Kranz, M. Kubillus, R. Ramakrishnan, O. A. von Lilienfeld, and M. Elstner, J. Chem. Theory Comput. **14**, 2341 (2018).
[150] J. Zhu, V. Q. Vuong, B. G. Sumpter, and S. Irle, MRS Commun. **9**, 867 (2019).
[151] A. W. Huran, C. Steigemann, T. Frauenheim, B. Aradi, and M. A. L. Marques, J. Chem. Theory Comput. **14**, 2947 (2018).
[152] N. Goldman, B. Aradi, R. K. Lindsey, and L. E. Fried, J. Chem. Theory Comput. **14**, 2652 (2018).
[153] H. Li, C. Collins, M. Tanha, G. J. Gordon, and D. J. Yaron, J. Chem. Theory Comput. **14**, 5764 (2018).
[154] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, J. Chem. Theory Comput. **11**, 2087 (2015).

© Author(s) 2020

[155]L. Shen, J. Wu, and W. Yang, J. Chem. Theory Comput. **12**, 4934 (2016).
[156]L. Shen and W. Yang, J. Chem. Theory Comput. **14**, 1442 (2018).
[157]V. W.-z. Yu, F. Corsetti, A. García, W. P. Huhn, M. Jacquelin, W. Jia, B. Lange, L. Lin, J. Lu, W. Mi, A. Seifitokaldani, Á. Vázquez-Mayagoitia, C. Yang, H. Yang, and V. Blum, Comput. Phys. Commun. **222**, 267 (2018).
[158]A. Marek, V. Blum, R. Johanni, V. Havu, B. Lang, T. Auckenthaler, A. Heinecke, H. J. Bungartz, and H. Lederer, J. Phys.: Condens. Matter **26**, 213201 (2014).
[159]T. Imamura, S. Yamada, and M. Machida, Prog. Nucl. Sci. Technol. **2**, 643 (2011).
[160]V. Hernandez, J. E. Roman, and V. Vidal, ACM Trans. Math. Software **31**, 351 (2005).
[161]J. Dongarra, M. Gates, A. Haidar, J. Kurzak, P. Luszczek, S. Tomov, and I. Yamazaki, *Numerical Computations with GPUs* (Springer, 2014), pp. 3–28.
[162]F. Corsetti, Comput. Phys. Commun. **185**, 873 (2014).
[163]L. Lin, M. Chen, C. Yang, and L. He, J. Phys.: Condens. Matter **25**, 295501 (2013).
[164]W. Dawson and T. Nakajima, Comput. Phys. Commun. **225**, 154 (2018).
[165]L. S. Blackford, J. Choi, A. Cleary, E. DAzevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet *et al.*, *ScaLAPACK Users' Guide* (SIAM, 1997).
[166]F. Tisseur and J. Dongarra, SIAM J. Comput. **20**, 2223 (1999).
[167]C. Vömel, ACM Trans. Math. Software **37**, 1 (2010).
[168]C. Bischof, X. Sun, and B. Lang, in *Proceedings of IEEE Scalable High Performance Computing Conference* (IEEE, 1994), pp. 23–27.
[169]K. Arturov, Intel Math Kernel Library (Intel MKL) 2018 update 2 ScaLAPACK symmetric eigensolver enhancements, https://software.intel.com/en-us/articles/intel-math-kernel-library-intel-mkl-2018-update-2-scalapack-symmetric-eigensolver; accessed 17 November 2019.
[170]A. M. N. Niklasson, Phys. Rev. B **66**, 155115 (2002).
[171]S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).
[172]C. F. A. Negre, S. M. Mnizsewski, M. J. Cawkwell, N. Bock, M. E. Wall, and A. M. N. Niklasson, J. Chem. Theory Comput. **12**, 3063 (2016).
[173]A. M. N. Niklasson, Phys. Rev. B **70**, 193102 (2004).
[174]A. M. Niklasson, S. M. Mniszewski, C. F. Negre, M. E. Wall, and M. J. Cawkwell, PROGRESS version 1.0, Technical Report No. LA-CC-16-068 (Los Alamos National Laboratory (LANL), Los Alamos, NM, USA, 2016).
[175]N. Bock, C. F. A. Negre, S. M. Mniszewski, J. Mohd-Yusof, B. Aradi, J.-L. Fattebert, D. Osei-Kuffuor, T. C. Germann, and A. M. N. Niklasson, J. Supercomput. **74**, 6201 (2018).
[176]N. Bock, S. Mniszewski, B. Aradi, M. Wall, C. F. A. Negre, and J. Mohd-Yusof, qmmd/bml v1.2.3, 2018.
[177]S. M. Mniszewski, M. J. Cawkwell, M. E. Wall, J. Mohd-Yusof, N. Bock, T. C. Germann, and A. M. N. Niklasson, J. Chem. Theory Comput. **11**, 4644 (2015).
[178]S. Tomov, J. Dongarra, and M. Baboulin, Parallel Comput. **36**, 232 (2010), part of special issue: Parallel Matrix Algorithms and Applications.
[179]A. Togo and I. Tanaka, Scr. Metall. **108**, 1 (2015).
[180]A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, J. Phys.: Condens. Matter **29**, 273002 (2017).
[181]M. Ceriotti, J. More, and D. E. Manolopoulos, Comput. Phys. Commun. **185**, 1019 (2014).
[182]M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, SoftwareX **1-2**, 19 (2015).
[183]See https://github.com/tomaskubar/gromacs-dftbplus for Gromacs repository fork; accessed 15 December 2019.
[184]T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[185]I. T. Todorov, W. Smith, K. Trachenko, and M. T. Dove, J. Mater. Chem. **16**, 1911 (2006).
[186]A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U. S. A. **99**, 12562 (2002).
[187]L. Alessandro and L. G. Francesco, Rep. Prog. Phys. **71**, 126601 (2008).
[188]A. Barducci, M. Bonomi, and M. Parrinello, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **1**, 826 (2011).
[189]G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[190]A. Barducci, M. Bonomi, and M. Parrinello, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **1**, 826 (2011).
[191]G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, Comput. Phys. Commun. **185**, 604 (2014).
[192]I. Mitchell, B. Aradi, and A. J. Page, J. Comput. Chem. **39**, 2452 (2018).
[193]See https://www.3dsbiovia.com/products/collaborative-science/biovia-materials-studio/ for BIOVIA Materials Studio; accessed 16 December 2019.
[194]See https://www.3dsbiovia.com/products/collaborative-science/biovia-materials-studio/materials-science-collections-for-pipeline-pilot.html for BIOVIA Materials Studio collection; accessed 15 December 2019.
[195]L. Guo, C. Qi, X. Zheng, R. Zhang, X. Shen, and S. Kaya, RSC Adv. **7**, 29042 (2017).
[196]K. Zhang, S. Yu, B. Jv, and W. Zheng, Phys. Chem. Chem. Phys. **18**, 28418 (2016).
[197]B. Delley, J. Chem. Phys. **92**, 508 (1990).
[198]B. Delley, J. Chem. Phys. **113**, 7756 (2000).
[199]See https://github.com/saydx for The Scientific ArraY Data EXchange library; accessed 23 February 2020.
[200]See https://github.com/dftbplus/mpifx for MPIFX library; accessed 15 December 2019.
[201]See https://github.com/dftbplus/scalapackfx for SCALAPACKFX library; accessed 15 December 2019.
[202]See https://github.com/aradi/fypp for Fypp preprocessor; accessed 15 December 2019.
[203]See https://github.com/cp2k/cp2k for CP2K software package; accessed 15 December 2019.

© Author(s) 2020